It Starts with **iGaze**: Visual Attention Driven Networking with Smart Glasses

Lan Zhang, Xiang-Yang Li, Wenchao Huang, Kebin Liu, Shuwei Zong,
Xuesi Jian, Puchun Feng, Taeho Jung, Yunhao Liu

Tsinghua University, China
Illinois Institute of Technology, USA
University of Science and Technology of China
Suzhou Institute for Advanced Study, China

“You are beautiful”

You're beautiful, you're beautiful,
You're beautiful, it's true.
I saw your face in a crowded place,
And I don't know what to do,
'Cause I'll never be with you....

It Starts with **iGaze**

You're beautiful, you're beautiful,

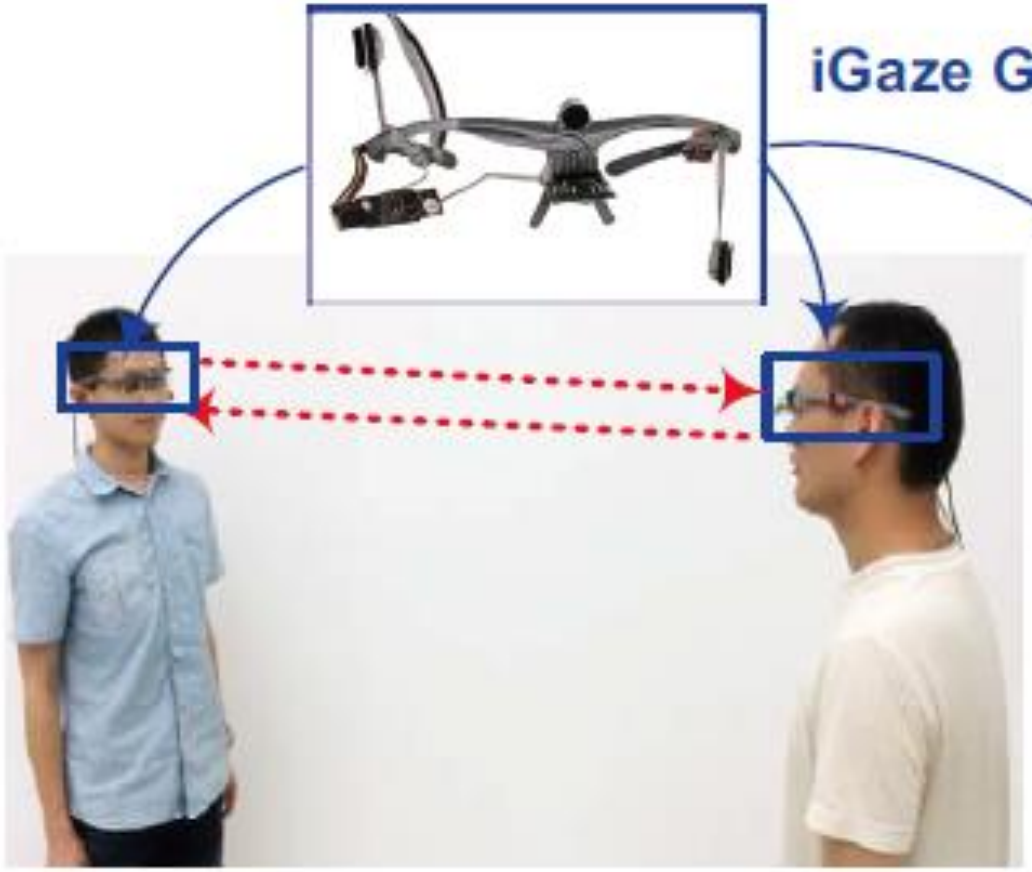
You're beautiful, it's true.

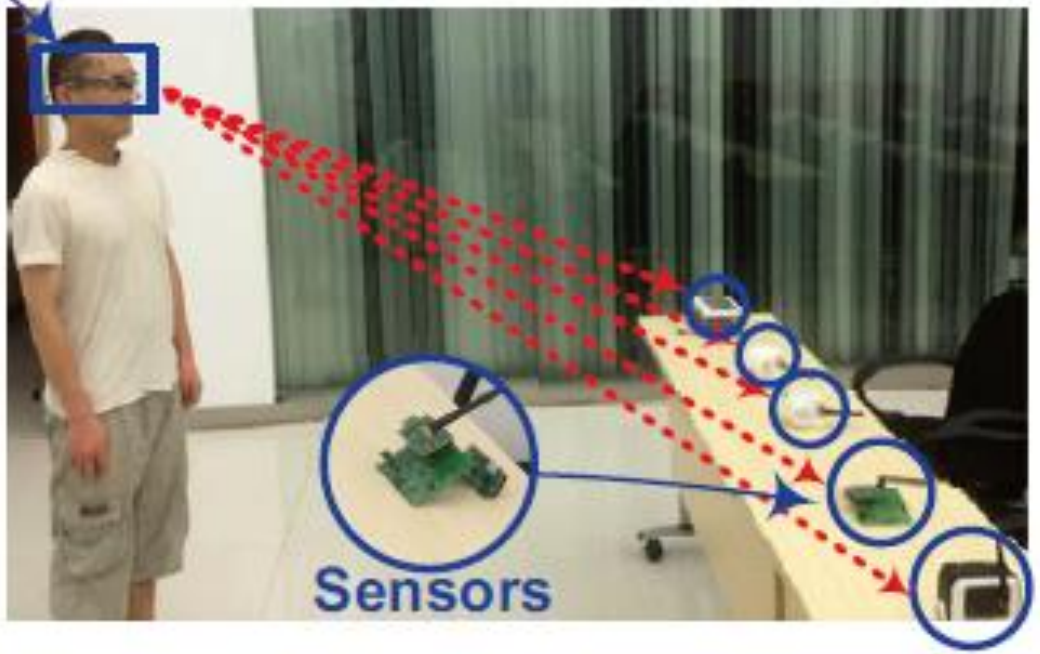
I saw your face in a crowded place,

And I start an eye gaze with you,

I'll be a friend of you.

Our **iGaze** Glasses

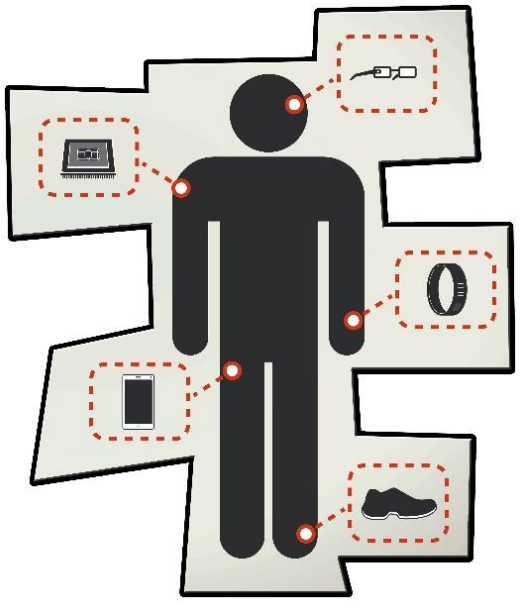




Outline

- Background
- Challenges
- System Design
- Implementation
- Evaluation
- Future Work

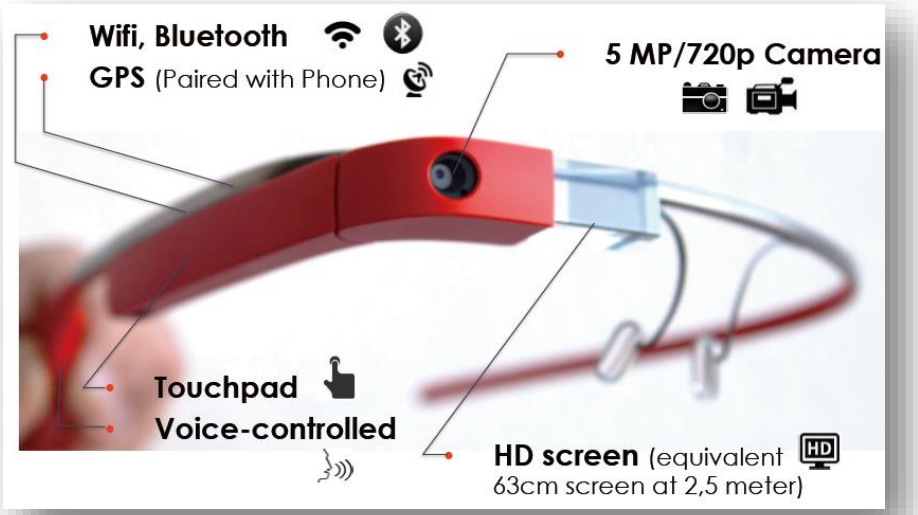
Next Biggest Thing: Wearable



2016, **170 million** wearable devices

2018:
30 billion $ market

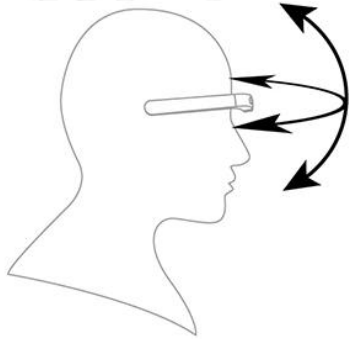
Smart Glasses



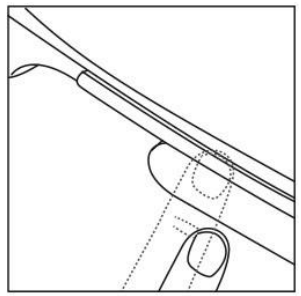
- **Cameras**: world/eye camera
- **Processing and Storage Units**
- **Sensors**: GPS, gyroscope, accelerometer
- **Networking modules:** Wi-Fi, Bluetooth

User interactions

Voice commands



Head gesture control



Hand control

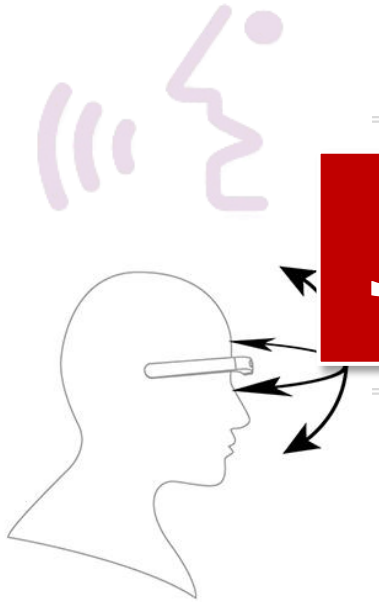
Functionalities

- Photo and video capture
- Social status updates
- Navigation

- Augmented reality overlay

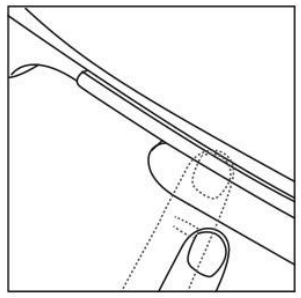
Existing Applications

User interactions



Voice commands

Head gesture control



Hand control

Functionalities

Just head-mounted smart phone?

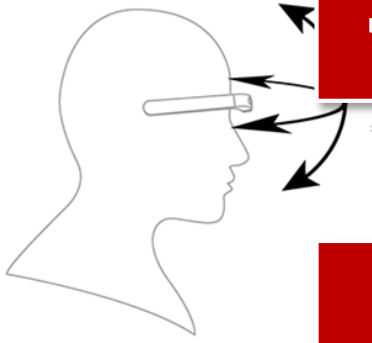
- Photo and video capture
- Social status updates
- Navigation
- …

Existing Applications

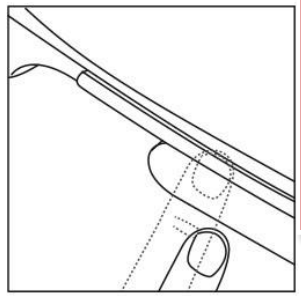
User interactions

Functionalities

Voice commands



Head gesture control



capture

ates

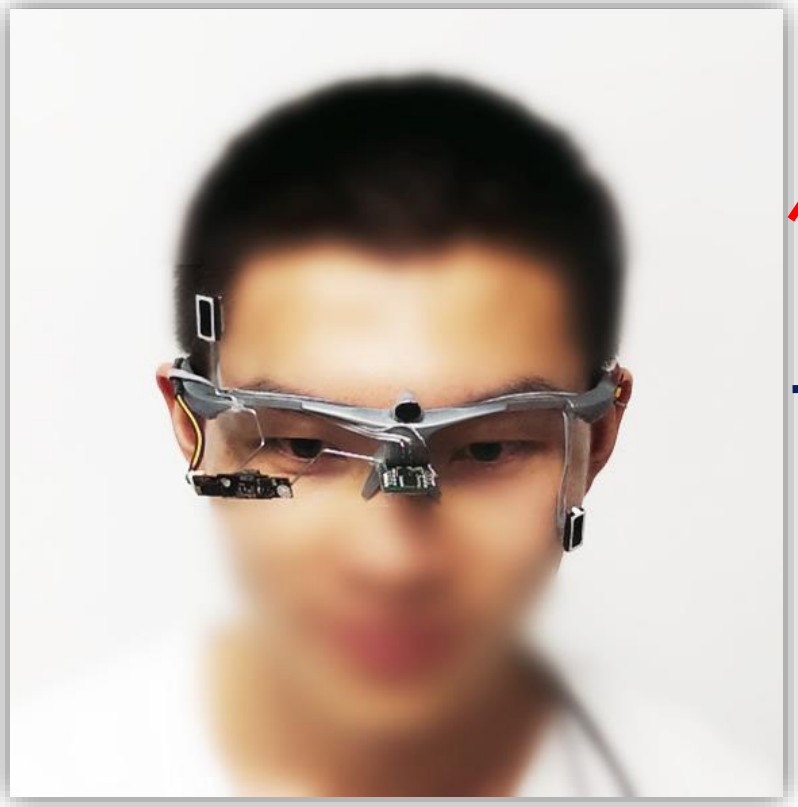
...

Just head-mounted smart phone?

Smart glasses are still looking for revolutionary applications!

Existing Applications

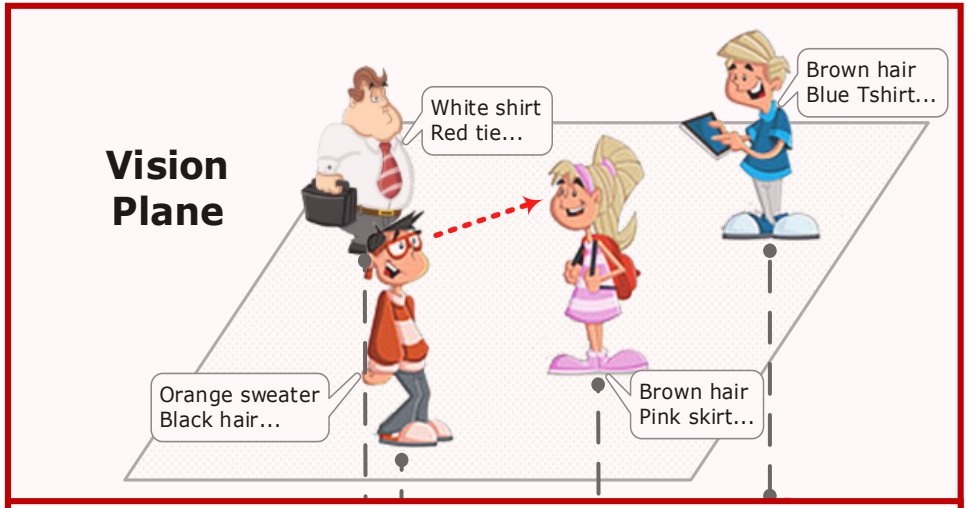
iGaze:
New Networking Mode → New Applications



- **Can** understand the user's visual attention using an eye camera
- **Can** automatically connect to the target of interest.
- **Can** run on top of existing networking protocols, e.g., Wi-Fi.

Application Modes of iGaze

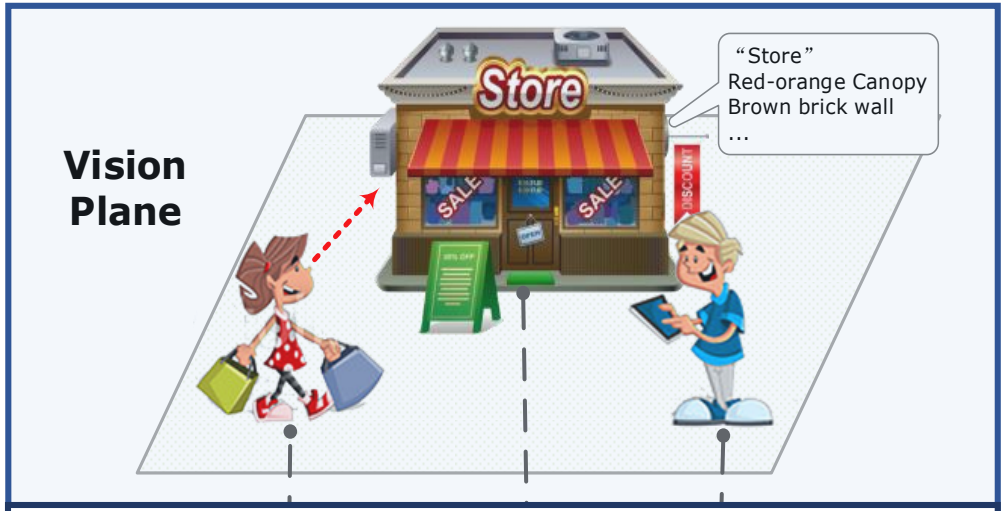
Bidirectional



Both users wear smart glasses.

Social networking scenario.

Unidirectional



One user wears smart glasses.

Object-oriented augment reality.

Potential Applications

Social

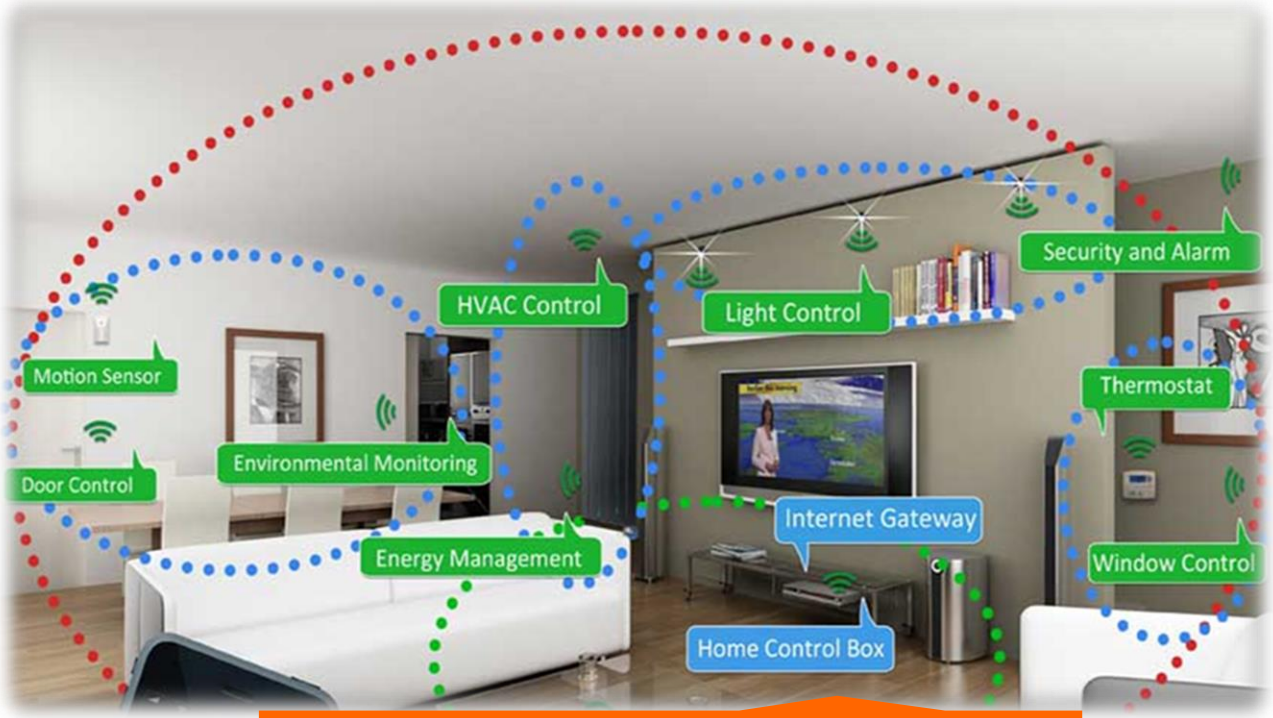
People can exchange information such as interests by simply looking at each other.

Potential Applications

Museum

Visitors can obtain the detailed descriptions of art works by glazing at it instead.

Potential Applications



Device Control

A house owner can turn on a smart appliance only by taking a glance at them.

Potential Applications

Advertising

When people look at a signboard or product, relevant promotion information can be sent to the potential customer immediately.

Targeted ads →big impulse buying market

Image recognition

QR Code

Scanning distance < 3m
Scanning angle <30 °

Recognition

Extensive training
High computation cost
Vulnerable to environment noise

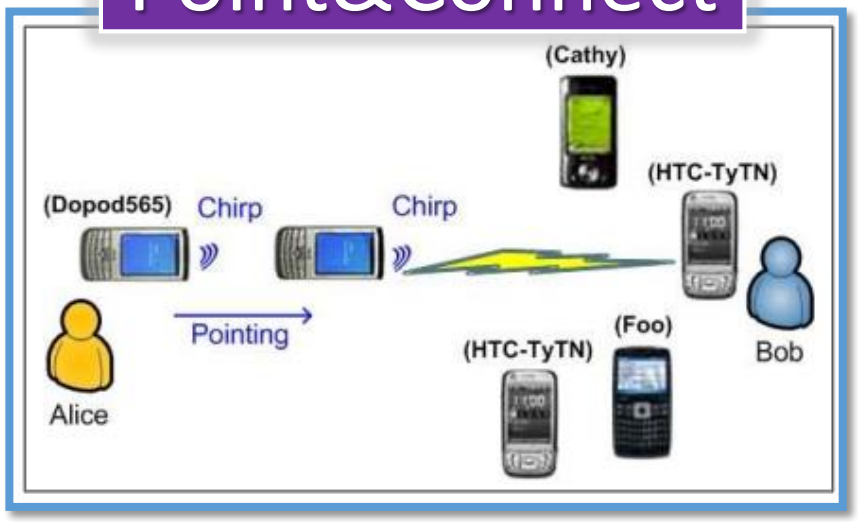
Existing Methods

Image recognition

Hand gesture based

[MobiSys 09’]

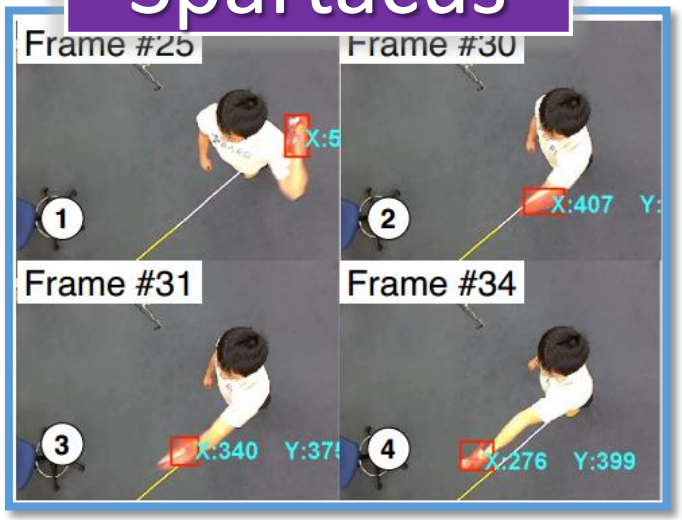
Point&Connect



Ranging based.

Displacement > 20cm.

Spartacus



[MobiSys 13’]

FFT based.

Speed: 2~6 m/s.

Existing Methods

Image recognition

Hand gesture based

Point&Connect

Spartacus

Targeting handhold smart devices.
Hard to achieve by head mounted devices.

Ranging based.
Displacement > 20cm.

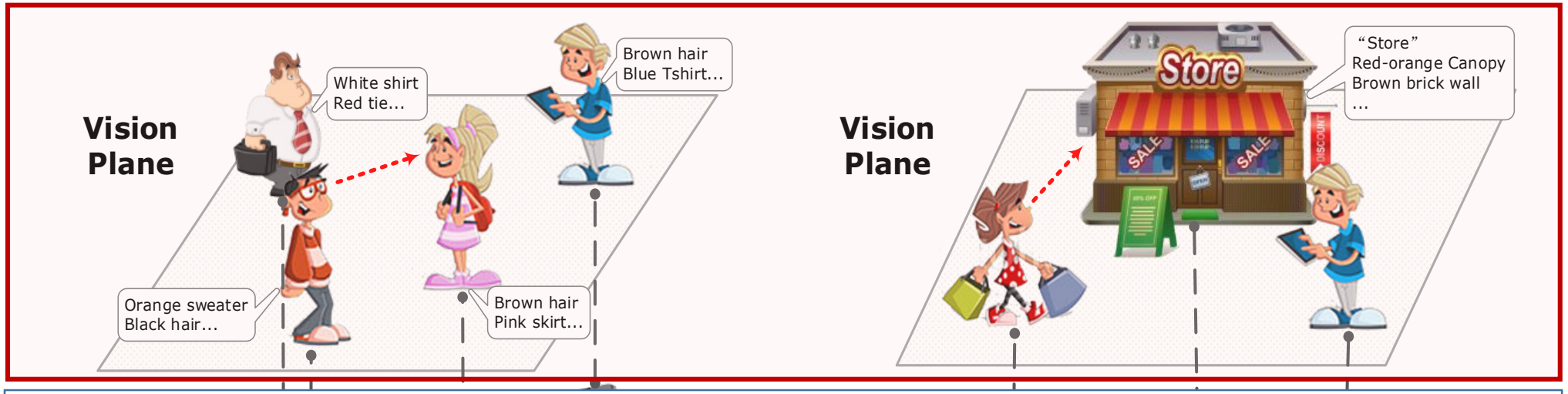
FFT based.
Speed: 2~6 m/s.

Existing Methods

Outline

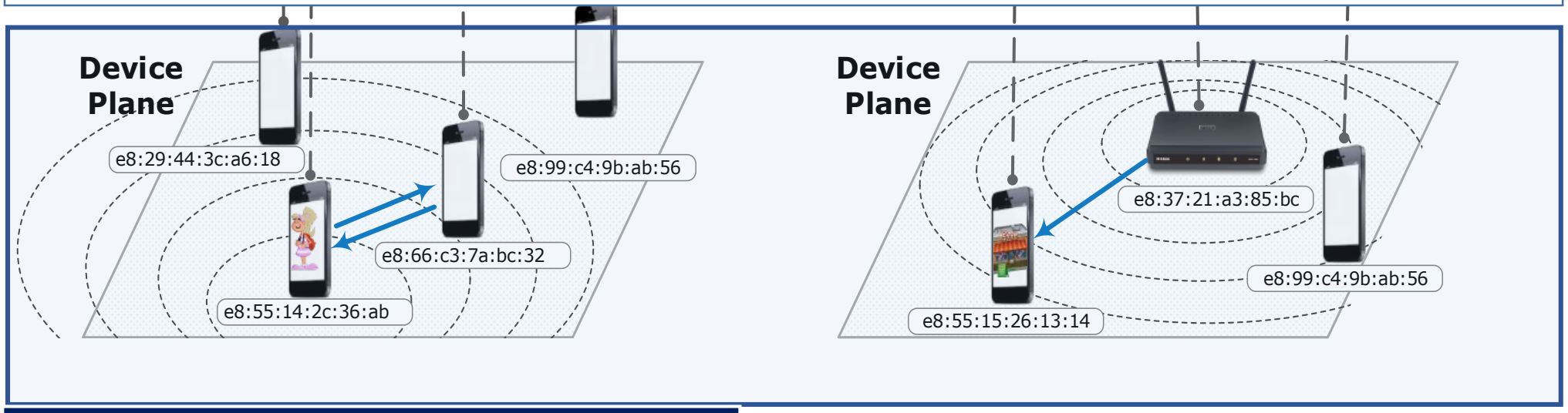
- Background
- **Challenges**
- System Design
- Implementation
- Evaluation
- Future Work

Vision Plane

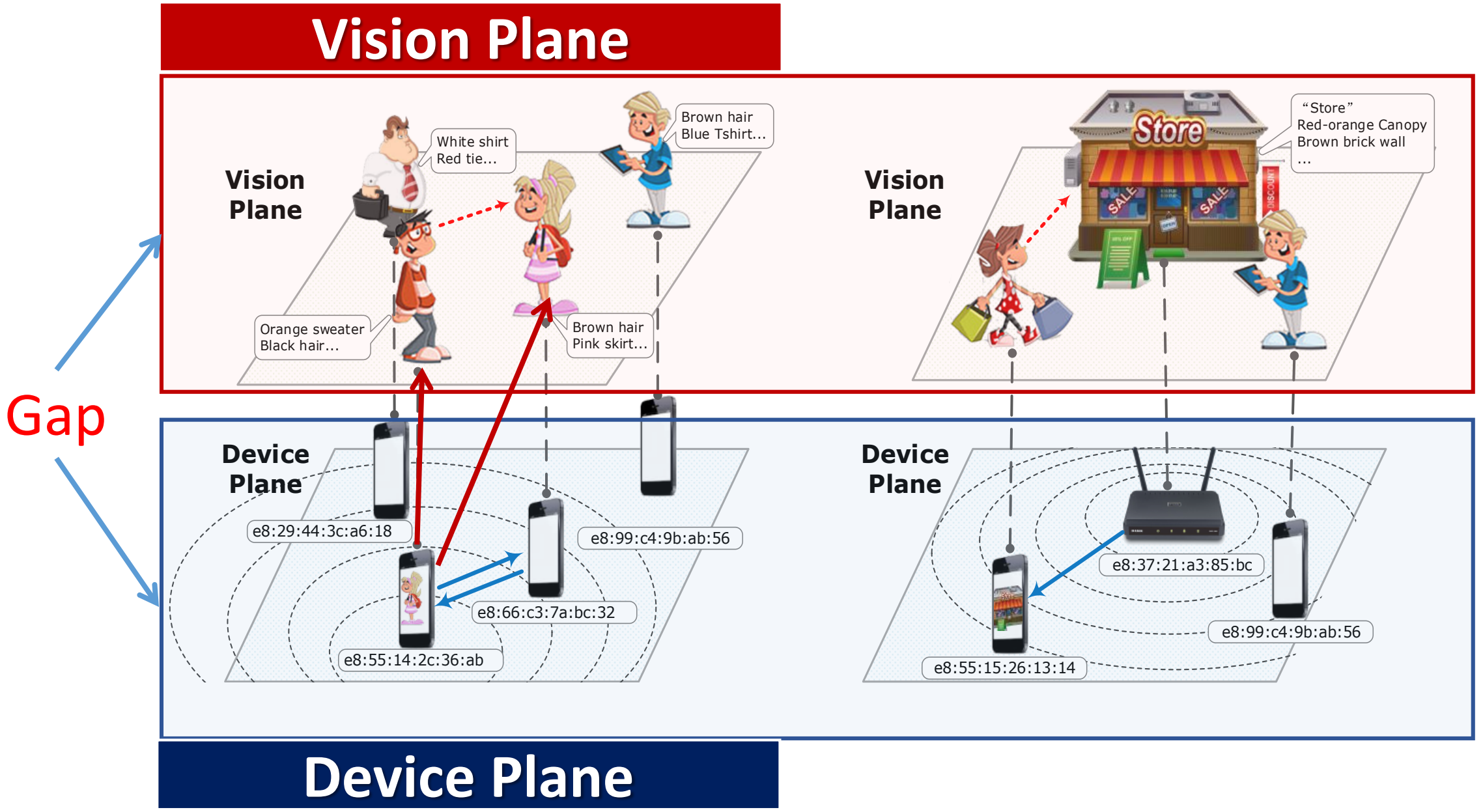


Device Plane

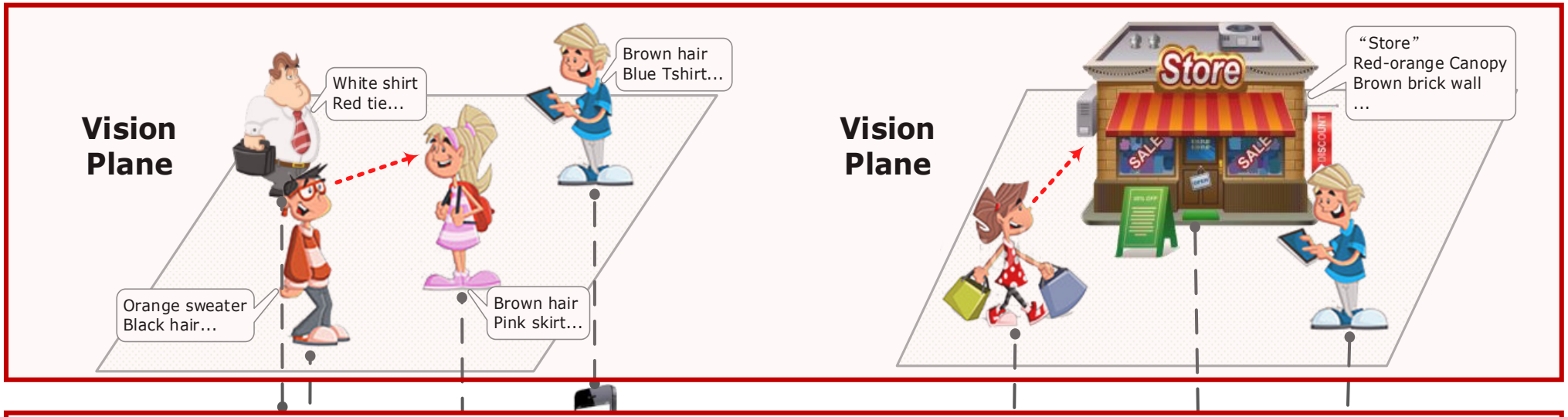
Vision Plane



Device Plane


Vision Plane
Vision Plane
White shirt
Red tie...
Brown hair
Blue Tshirt...
Orange sweater
Black hair...
Brown hair
Pink skirt...
Vision Plane
Store
"Store"
Red-orange Canopy
Brown brick wall
...
Gap
Device Plane
e8:29:44:3c:a6:18
e8:99:c4:9b:ab:56
e8:66:c3:7a:bc:32
e8:55:14:2c:36:ab
Device Plane
e8:37:21:a3:85:bc
e8:99:c4:9b:ab:56
e8:55:15:26:13:14
Device Plane

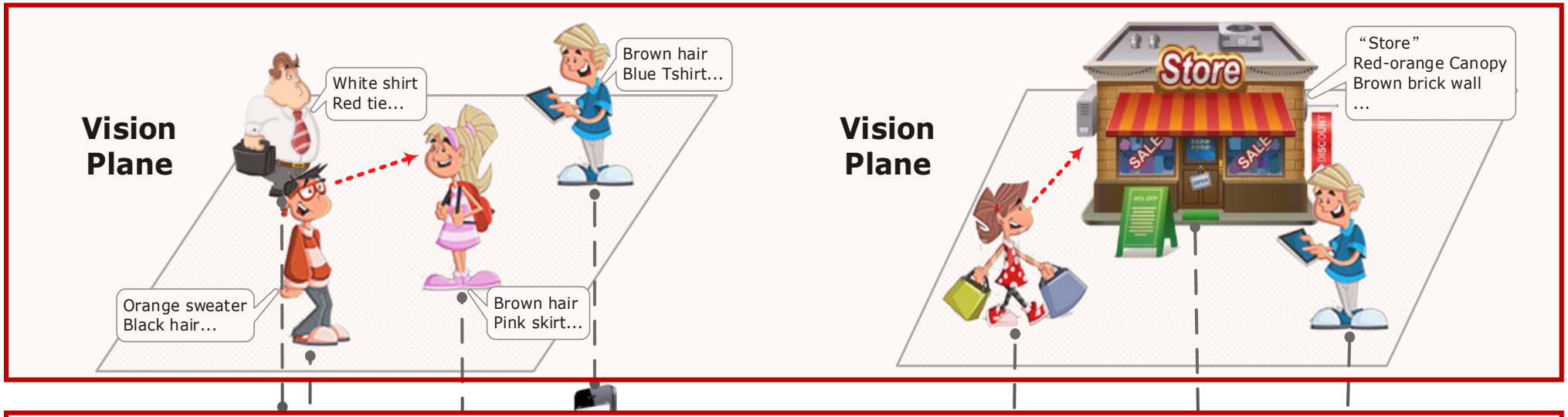
Vision Plane



To accurately capture the vision plane attention of a user in real time

Goal

Vision Plane



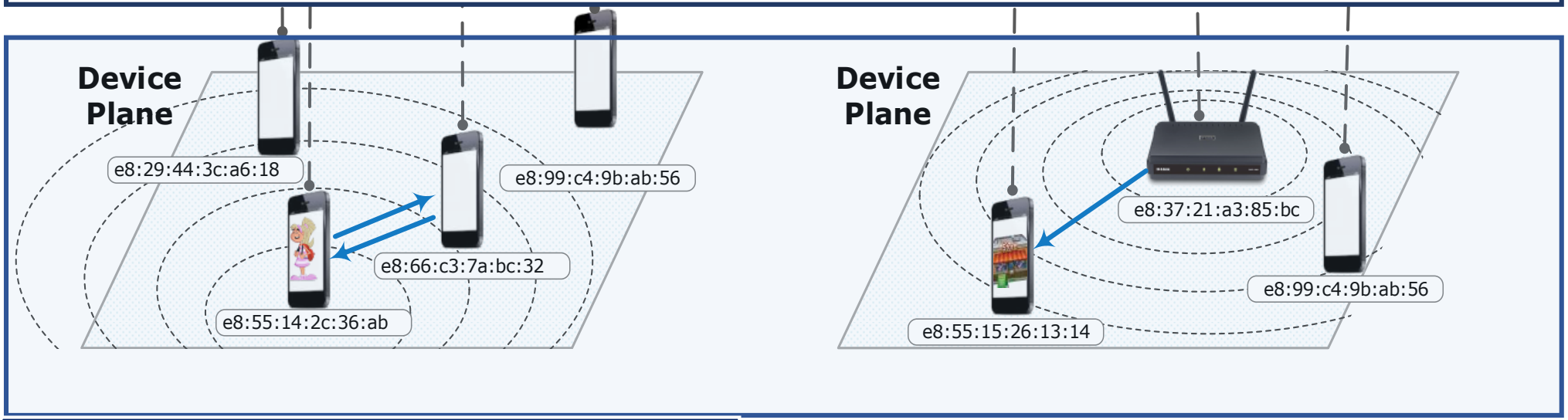
- One low-resolution camera.
- No training and manual calibration.
- No restriction on users' movement.
- Low computing and energy cost.

Challenge

Goal

Find another identifier of devices good for matching

Candidate: device direction



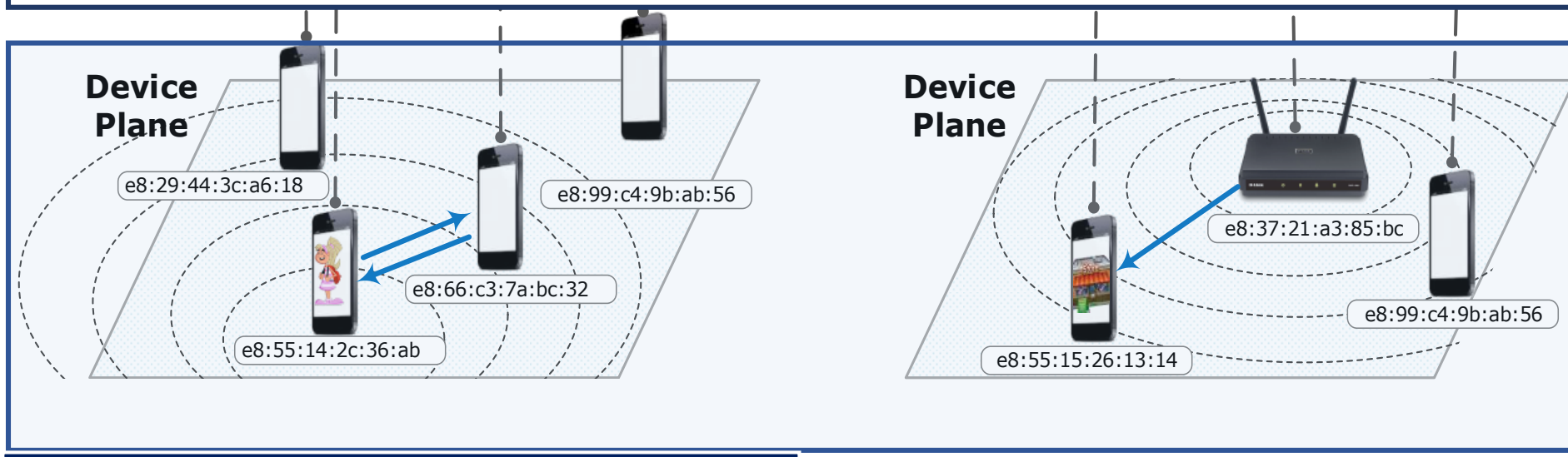
Device Plane

Challenge

Convenient user control

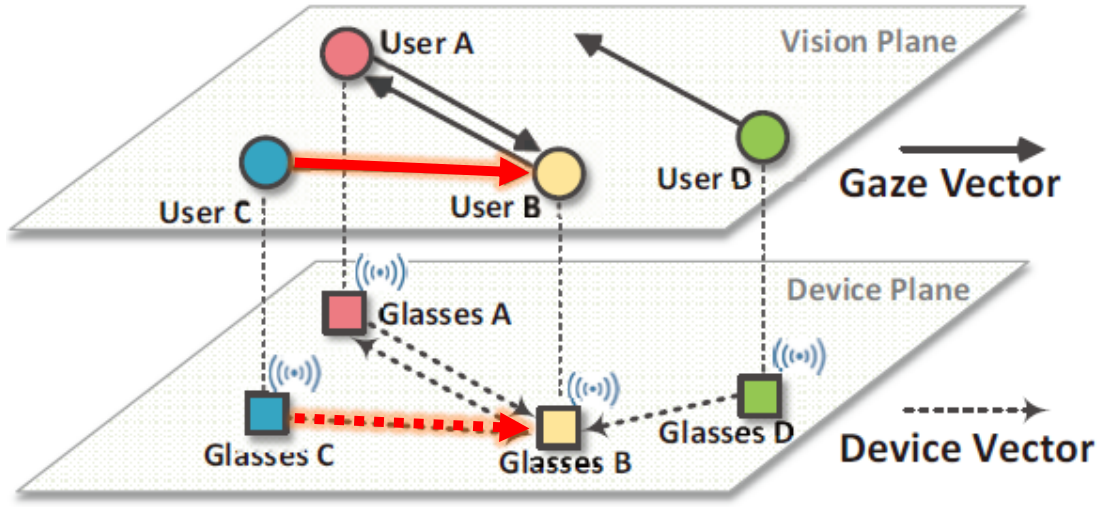
Accuracy, Delay, Energy cost

Easy 1-1 mapping with users in vision plane!



Device Plane

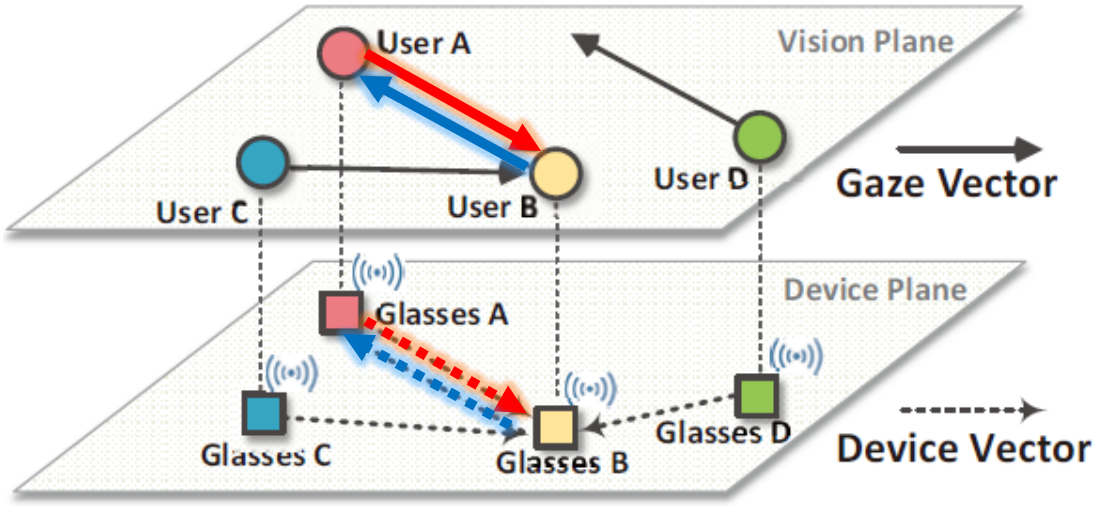
Principle of Gaze-based Networking



Observation 1

Given an observer and his/her gaze vector, only the device vector to the correct visual target's device is consistent with the gaze vector.

Principle of Gaze-based Networking



Observation 2

Given a pair of users who are looking at each other, their gaze vectors have opposite directions.

Solution

Vision Plane

- Develop a low-cost glasses hardware
  - embedded with our attention capture software,
  - capture the gaze direction using the orientation sensor and eye camera.

Device Plane

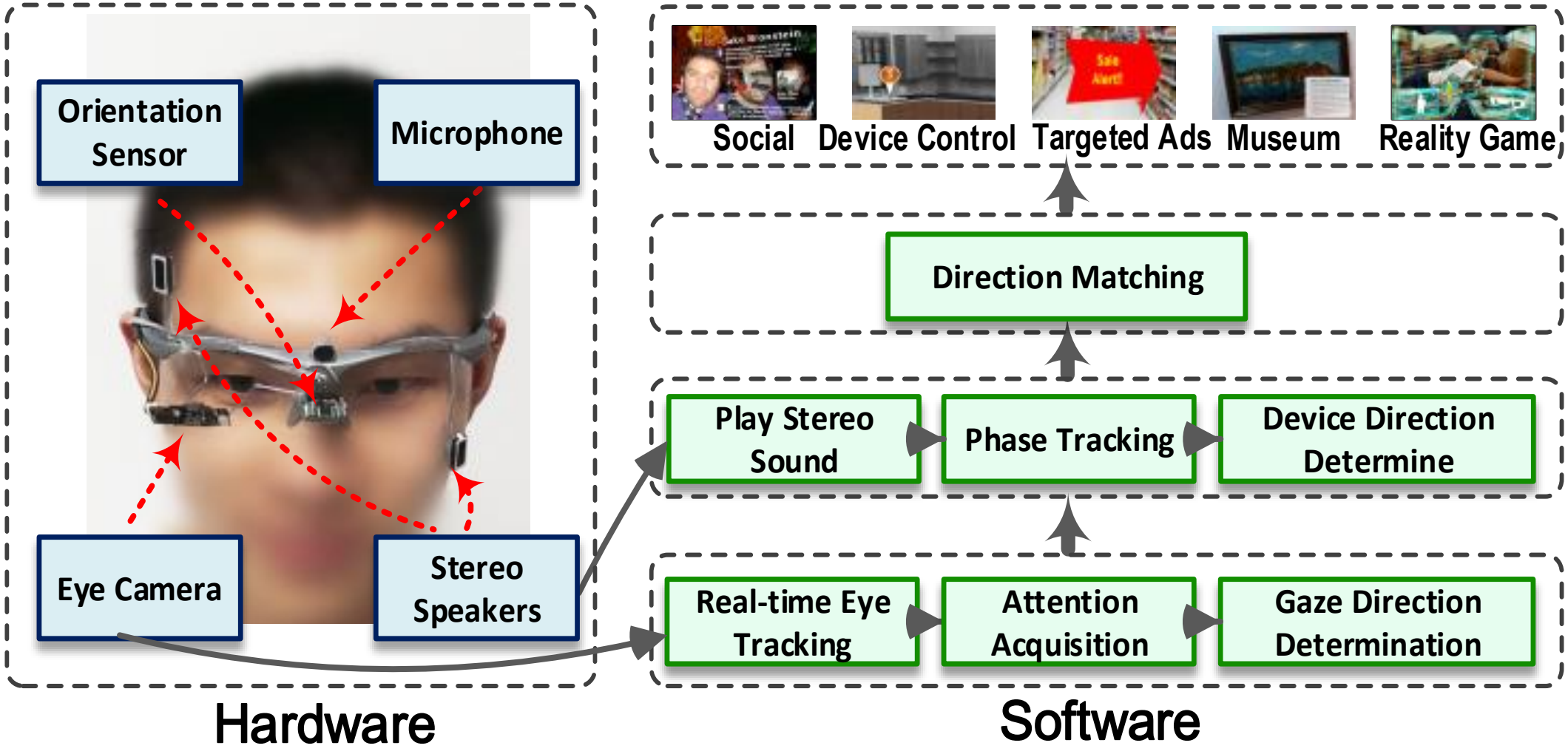
- Capture the direction between glasses
  - leveraging Doppler effect caused by arbitrary mild movement of head-mounted speakers.

Solution

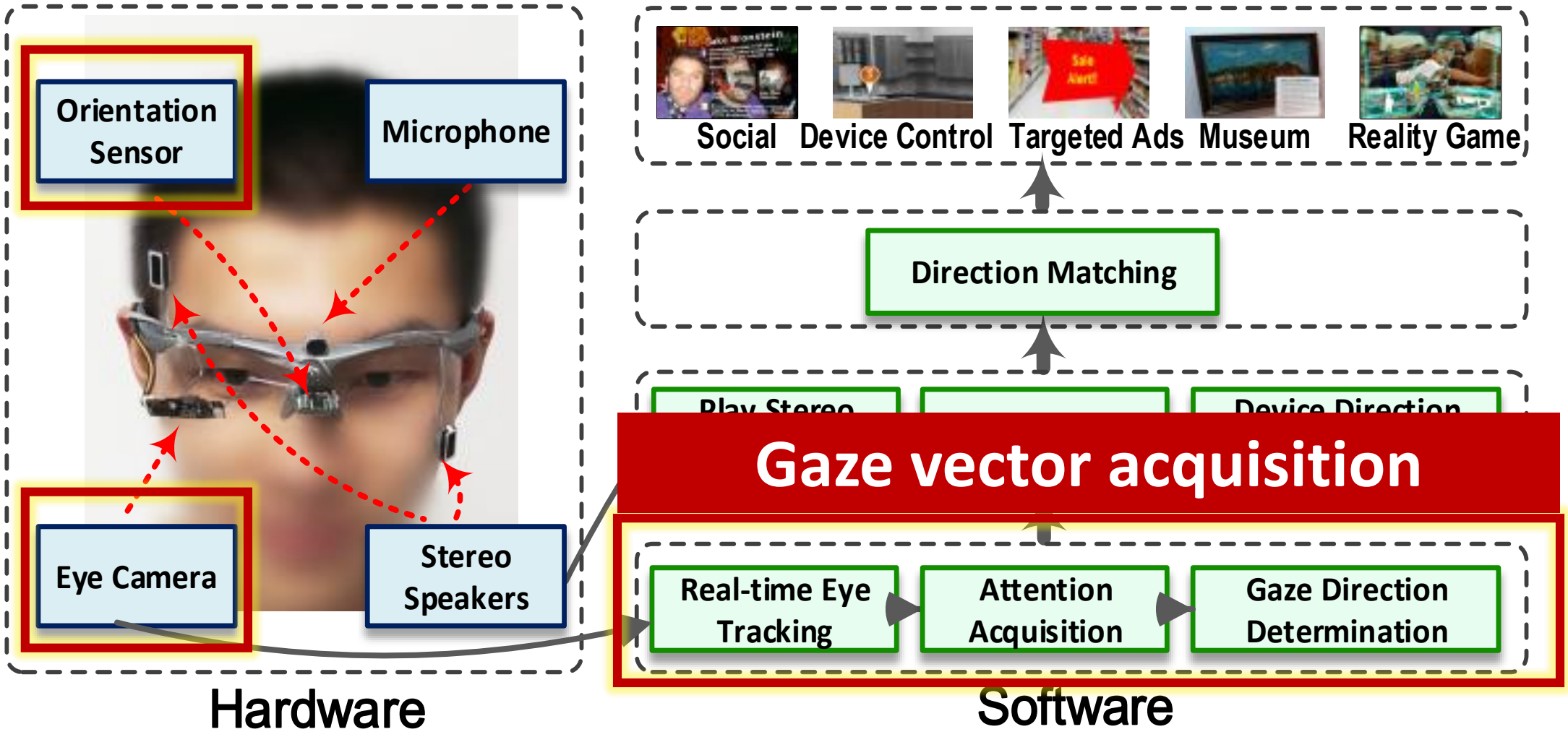
Outline

- Background
- Challenges
- **System Design**
- Implementation
- Evaluation
- Future Work

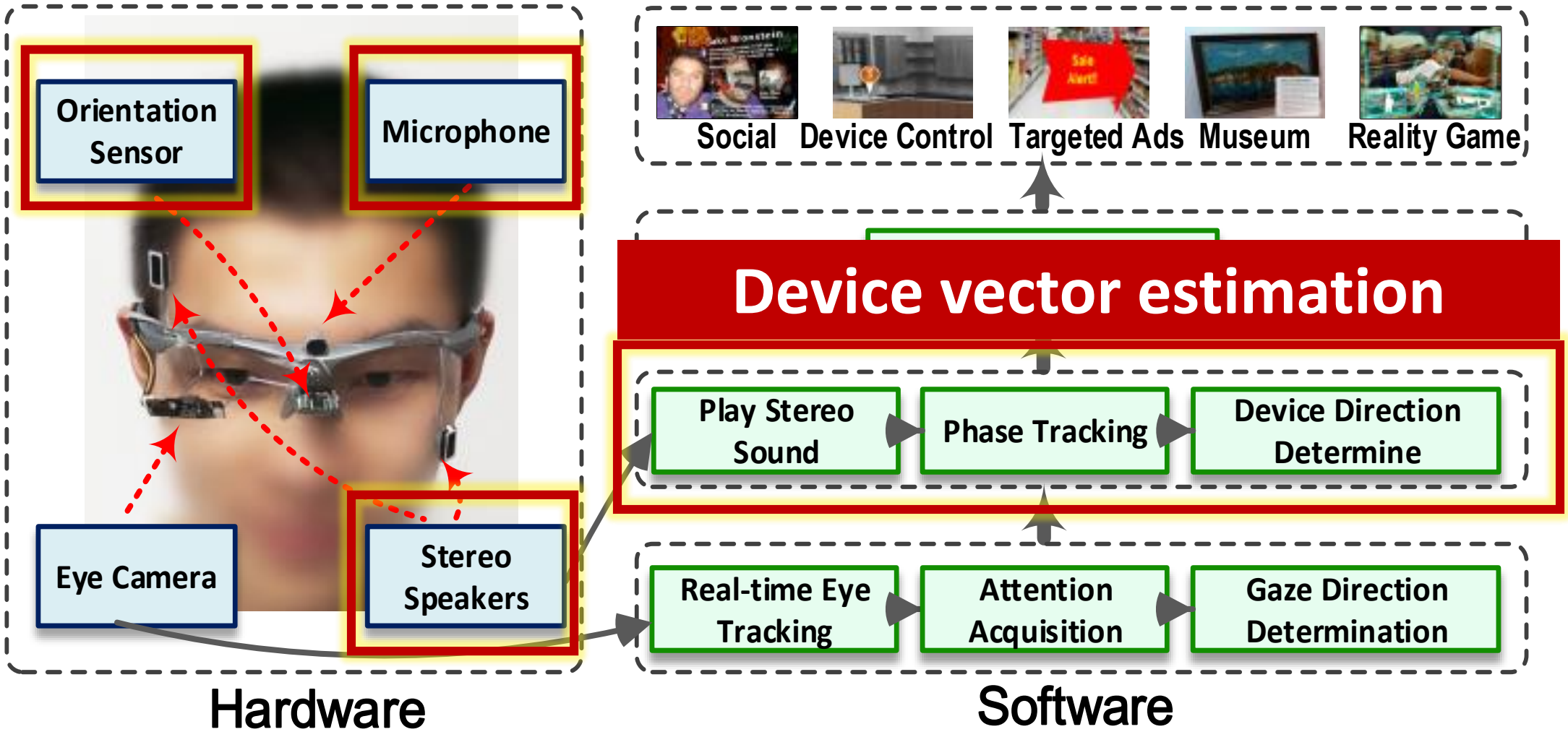
Hardware & Software Architecture



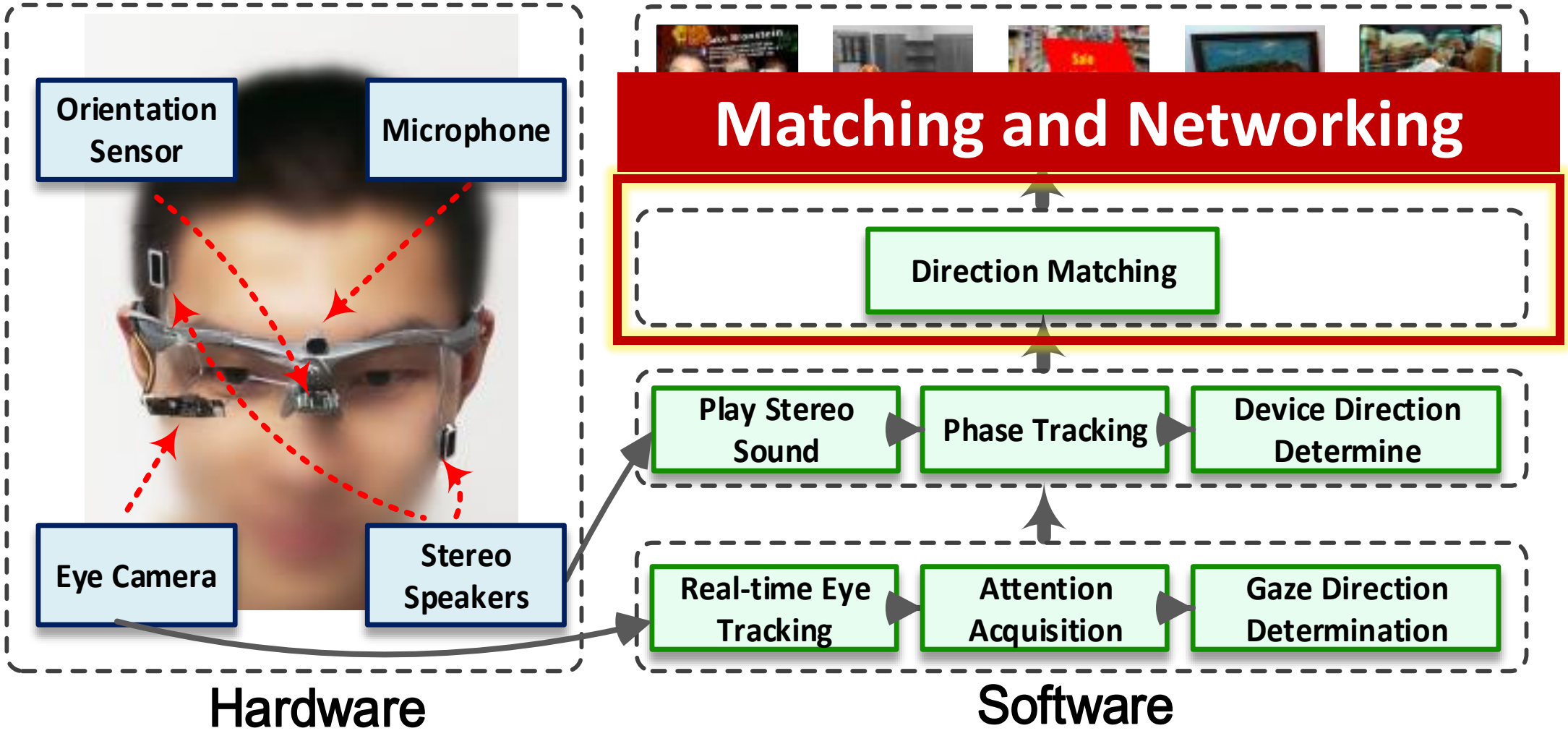
Hardware & Software Architecture



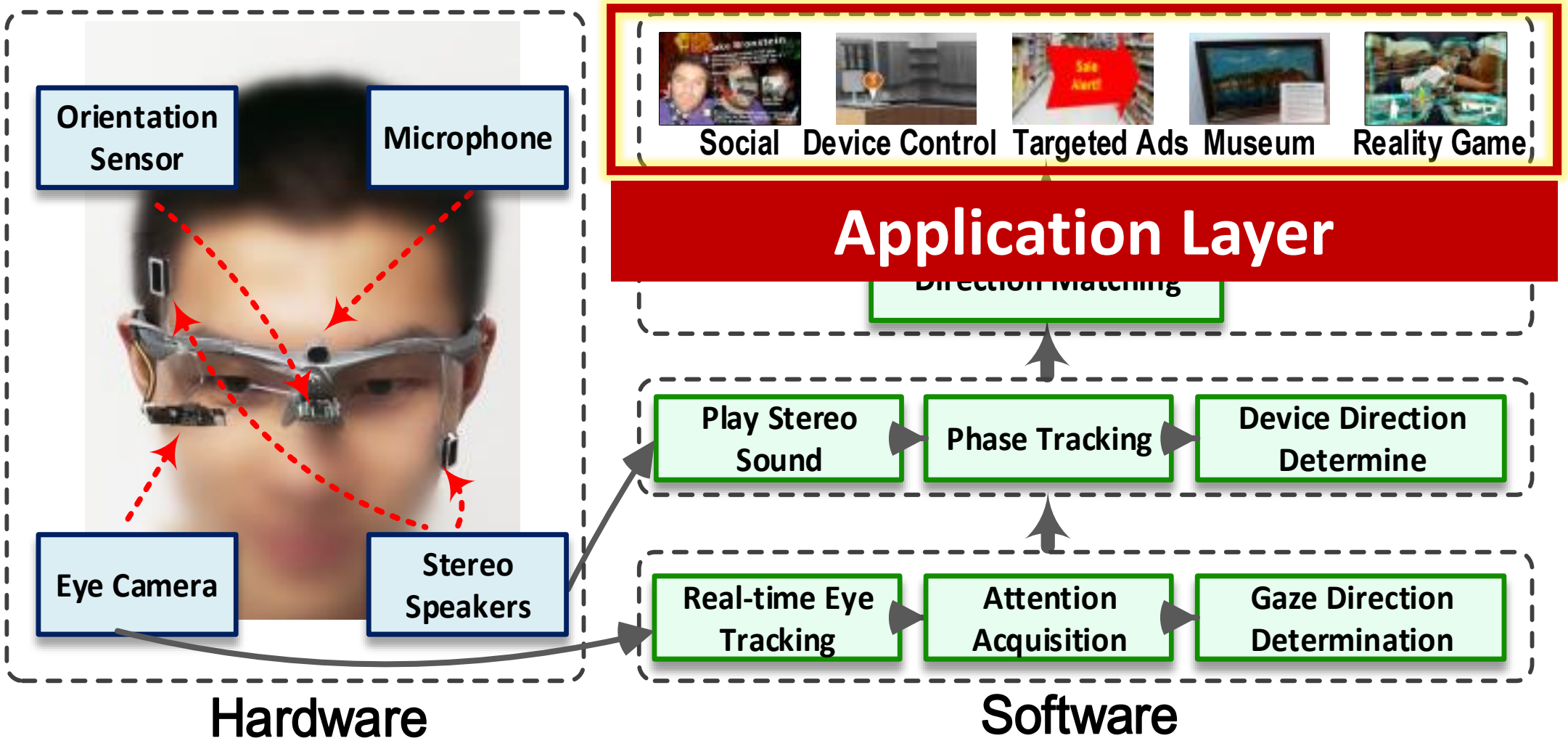
Hardware & Software Architecture



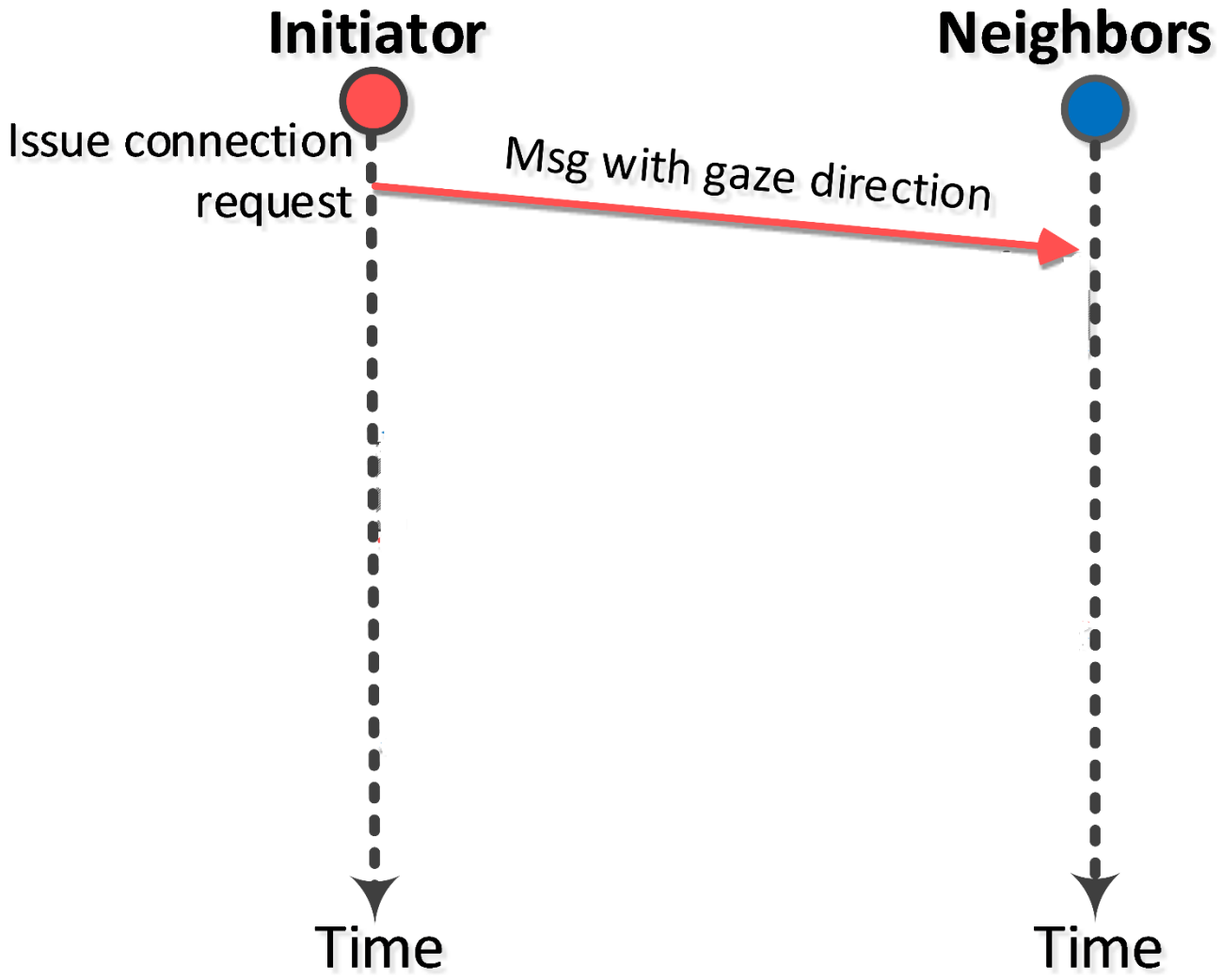
Hardware & Software Architecture



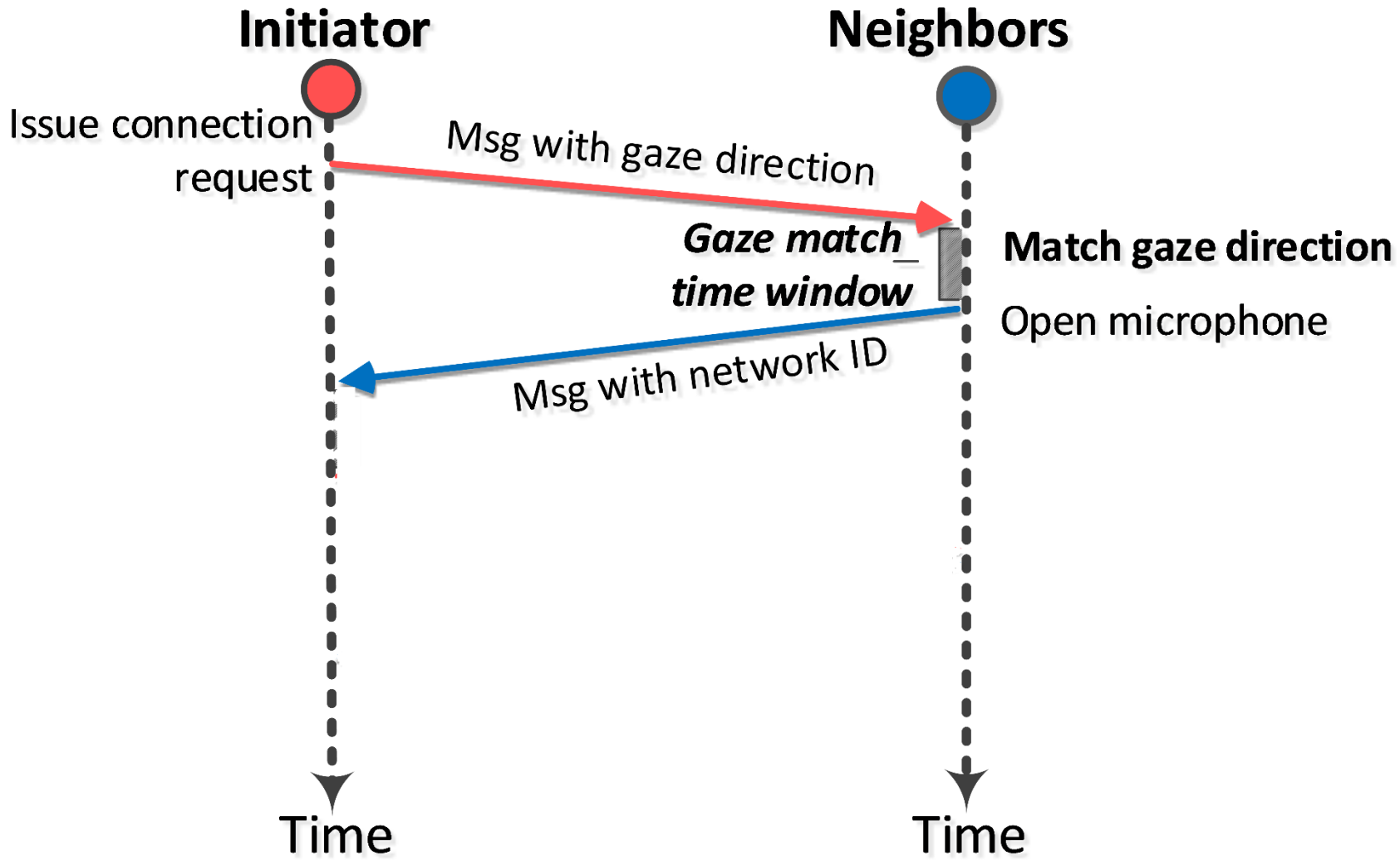
Hardware & Software Architecture



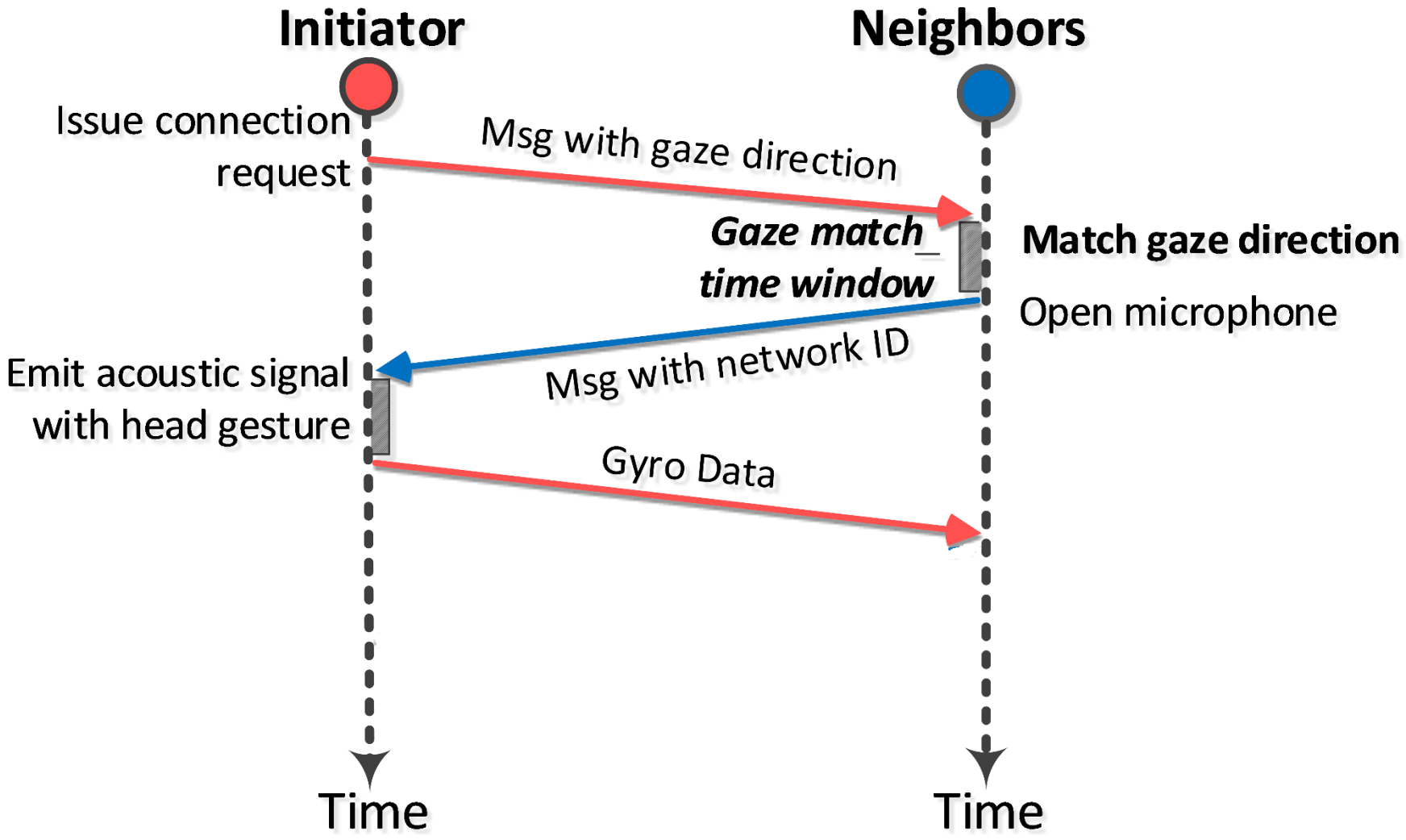
Visual Attention Networking (VAN) Protocol



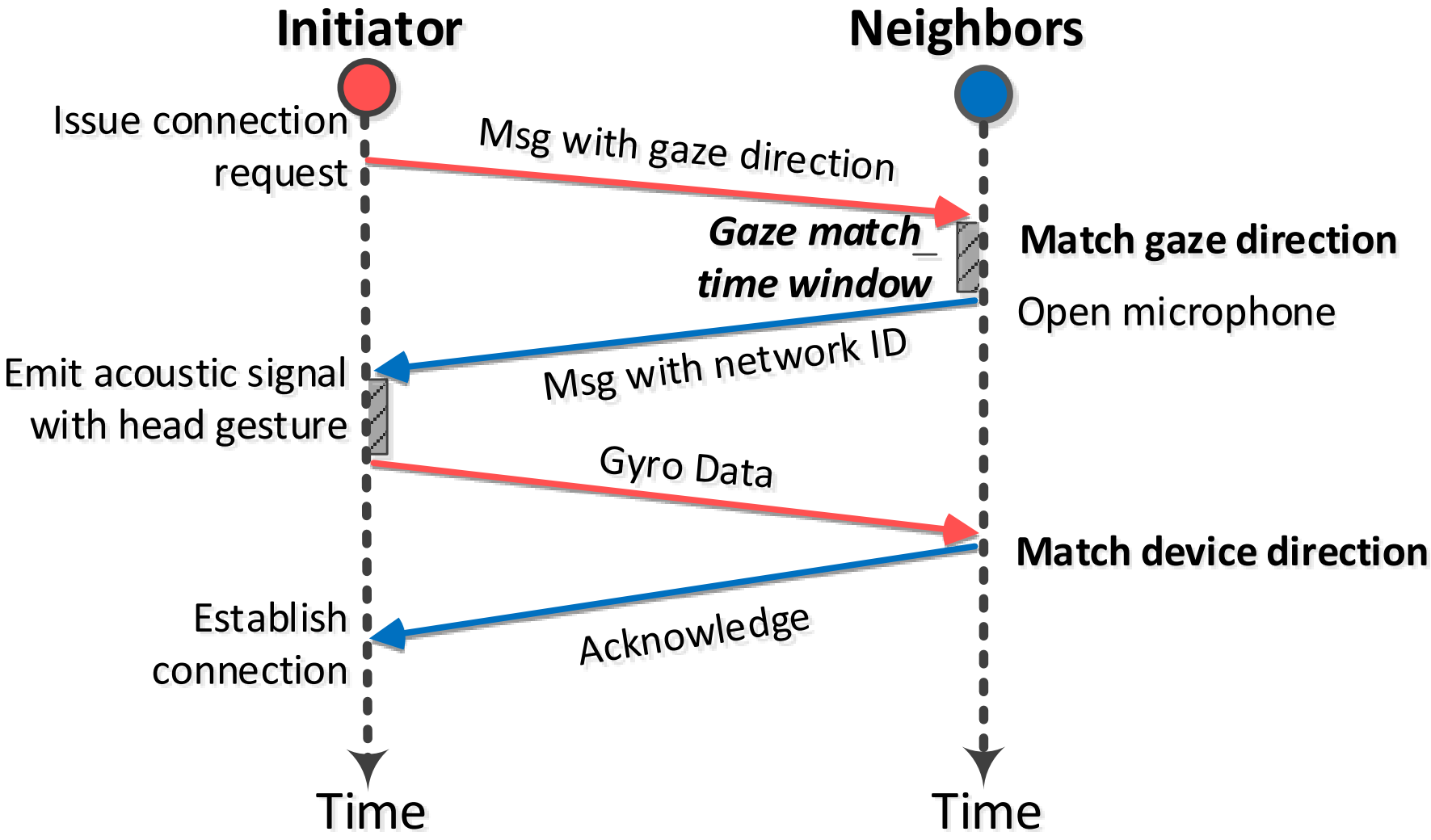
Networking Protocol



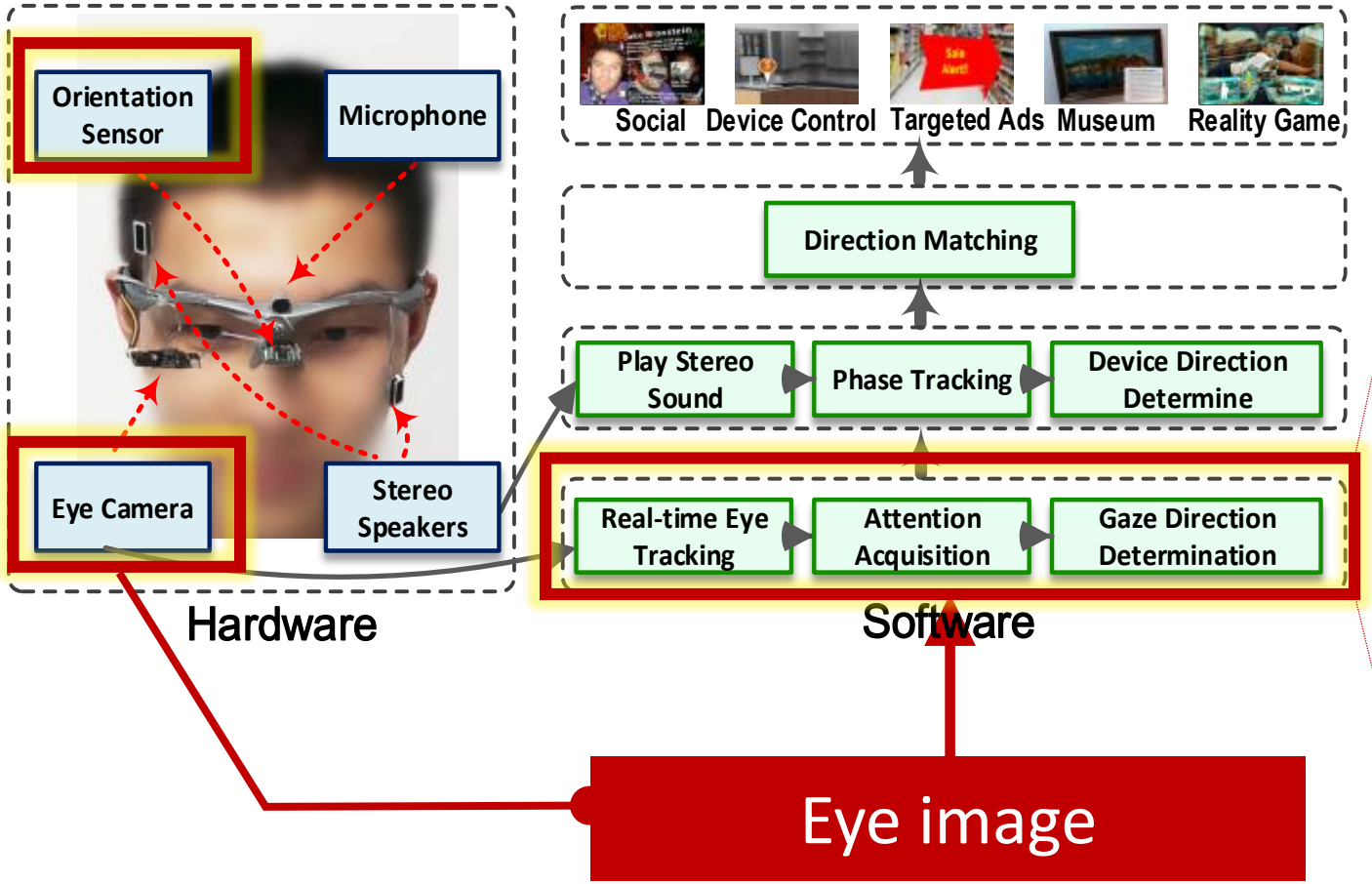
Networking Protocol



Networking Protocol



System Design

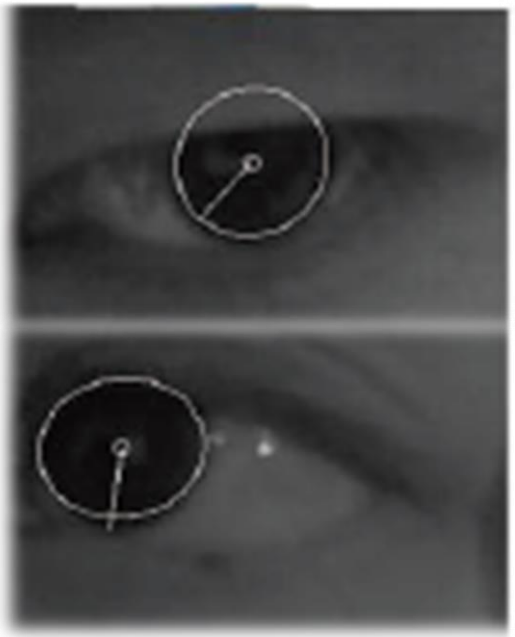


Gaze Vector Acquisition

- **Real-time eye tracking**: captures the movement of a user's eye.
- **Attention acquisition**: detect a visual attention when the gaze lasts for a reasonable time.
- **Gaze vector determination**: calculates the corresponding gaze vector to the visual target.

Visual Attention Detection

Real-time Eye Tracking

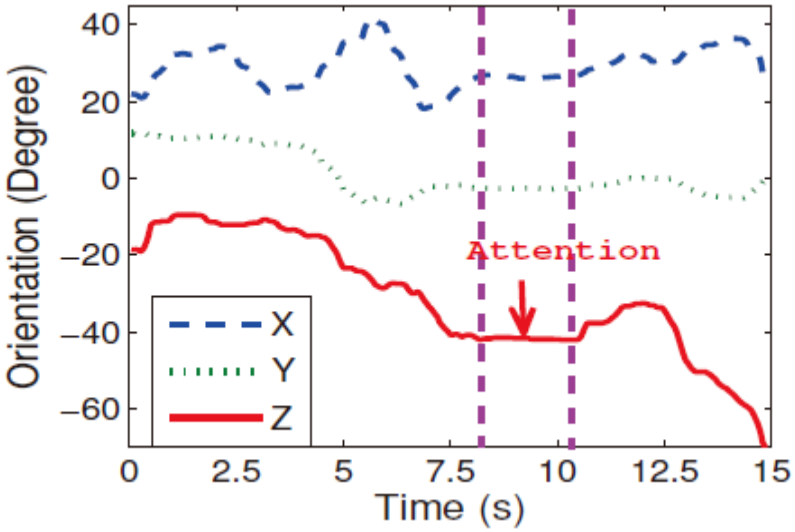


- Two basic eye movement statuses: **fixation** and **saccade**.
- visual attention ←-→ eye fixation.
- We use
  - an eye-movement velocity threshold to detect the fixation
  - a fixation duration window threshold for visual attention

Visual Attention Detection

On-demand Eye Tracking

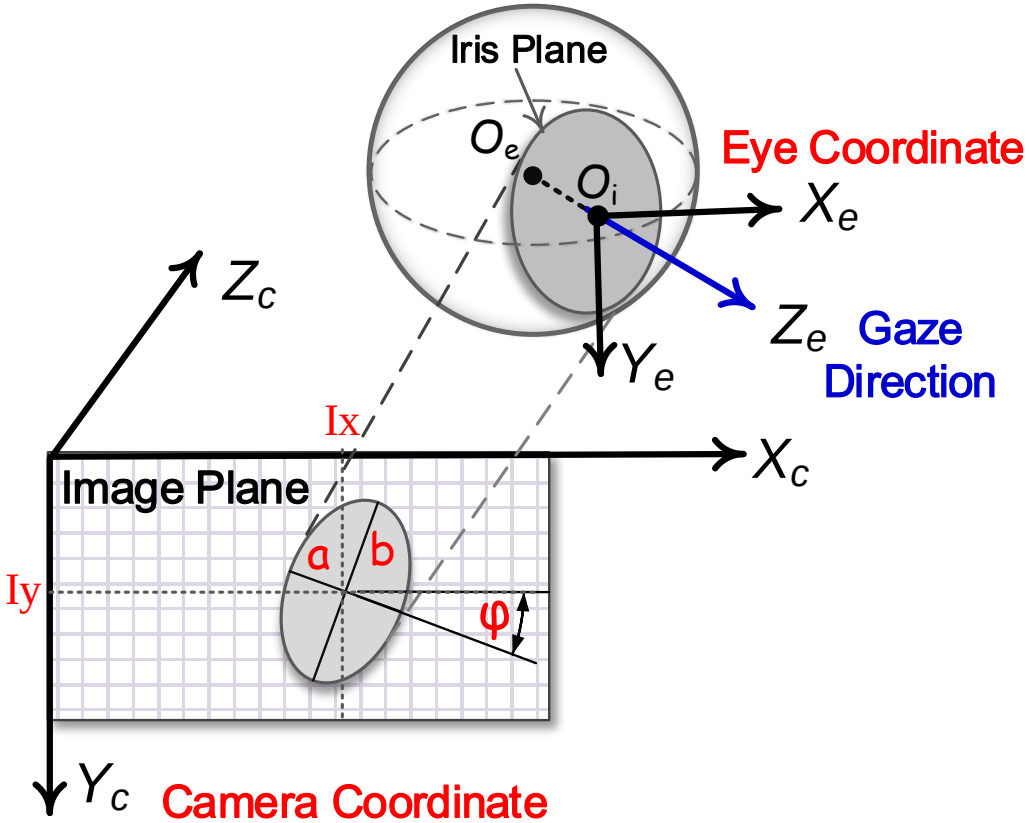
- Continuous image capturing and analysis could be power and computation intensive.
- To further reduce the cost, we detect head fixation using gyro data before invoking the eye camera.



Gyro data of head movement.

Determine Gaze Direction

Projection Model

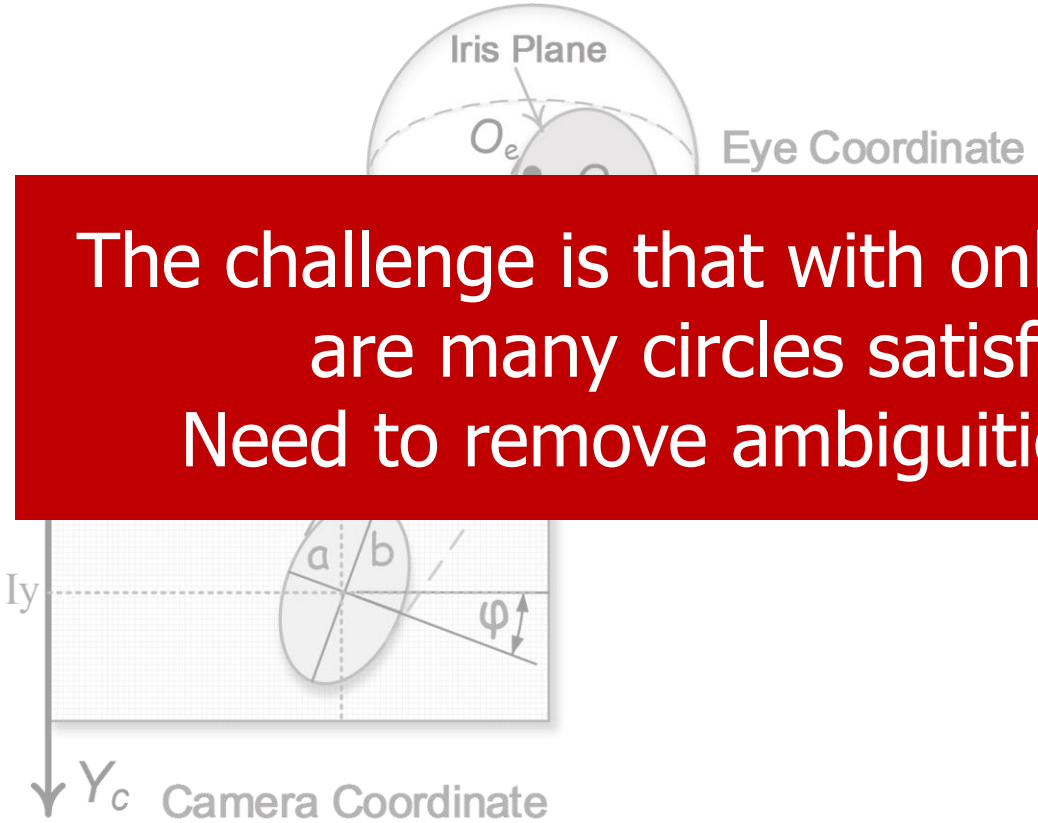


Basic Idea

- The iris contour is a circle, but its projection on the image plane is elliptical.
- When look straight ahead, the projection looks more like a circle;
- When look off to one side, it looks more close to an ellipse.
- We can estimate the pose of the iris circle by back-projecting the ellipse onto a circle in 3D space.

Determine Gaze Direction

Projection Model



Basic Idea

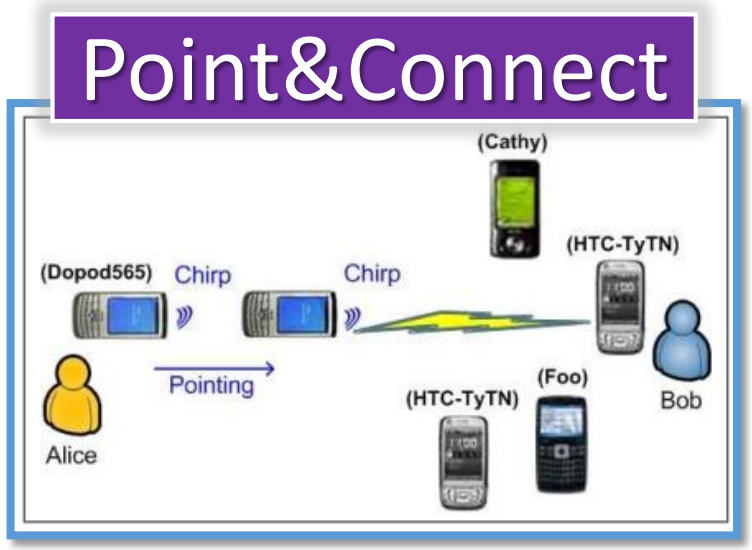
- The iris contour is a circle, but its projection on the image plane is elliptical.

The challenge is that with only a single elliptical image, there are many circles satisfying the projection cone.
Need to remove ambiguities caused by single camera!!

- We can estimate the pose of the iris circle by back-projecting the ellipse onto a circle in 3D space.

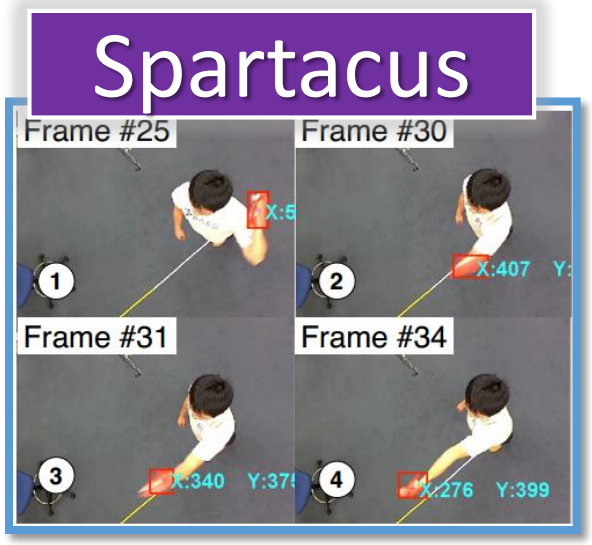
Estimate Device Vector

[MobiSys 09']



Ranging based.

Displacement > 20cm.



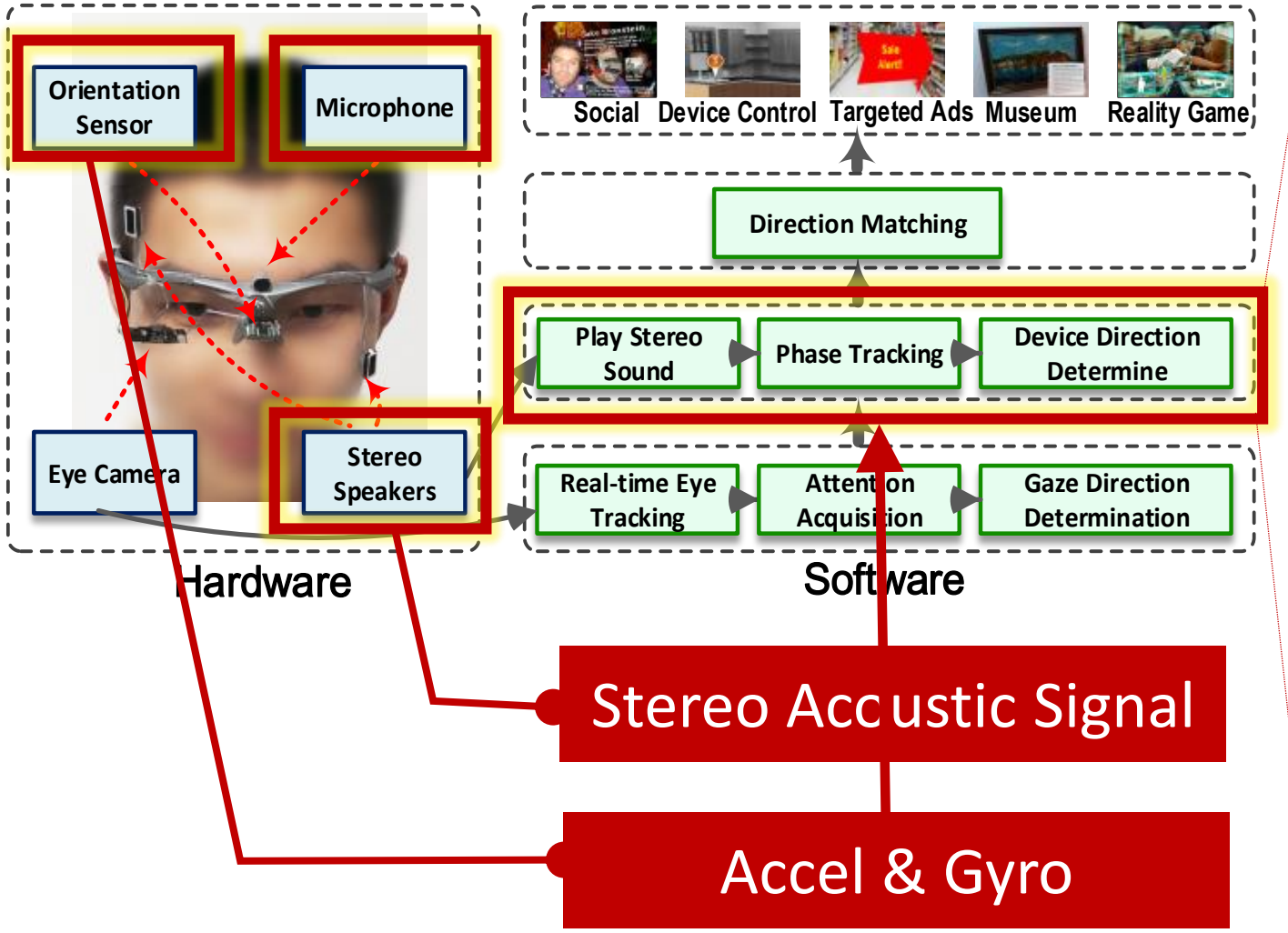
[MobiSys 13']

FFT based.

Speed: 2~6 m/s.

The main challenge is how to determine accurate device direction with only mild head gesture?

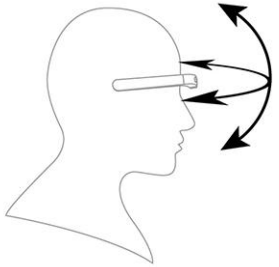
System Design



Device Vector Estimation

- Use two head-mounted speakers, based on the Phase Locked Loop (PLL) technique.
- The measuring accuracy of the relative displacement is less than 1mm.

Device Direction Estimation

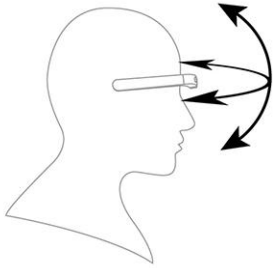


Make head gesture

Makes a head gesture (e.g., head nod, shake).

Two speakers play two sine inaudible waves at different frequencies.

Device Direction Estimation



$$r(t) = \sin(2\pi f_a t + \phi_t)$$

$$d = \|RA_1\| - \|RA_2\| = \frac{v_a}{2\pi f_a}(\phi_{t_1} - \phi_{t_2})$$

Make head gesture

Makes a head gesture (e.g., head node).

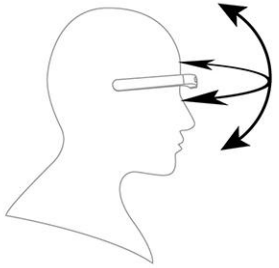
Two speakers play two sine inaudible waves at different frequencies.

Get relative displacement

The receiver tracks the precise phase of the received signal by PLL.

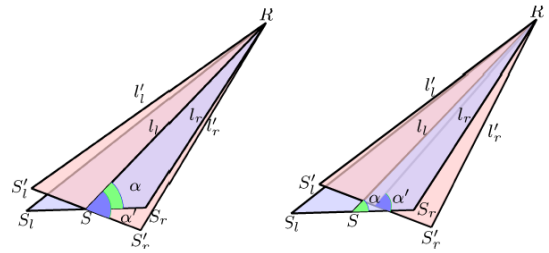
The phase shift is in proportion to relative displacement.

Device Direction Estimation



$$r(t) = \sin(2\pi f_a t + \phi_t)$$

$$d = \|RA_1\| - \|RA_2\| = \frac{v_a}{2\pi f_a}(\phi_{t_1} - \phi_{t_2})$$



(a) Case 1. (b) Case 2.

Make head gesture

Makes a head gesture (e.g., head node).

Two speakers play two sine inaudible waves at different frequencies.

Get relative displacement

The receiver tracks the precise phase of the received signal by PLL.

The phase shift is in proportion to relative displacement.

Get relative direction

Obtain the relative device direction via the relative displacements of two speakers.

Outline

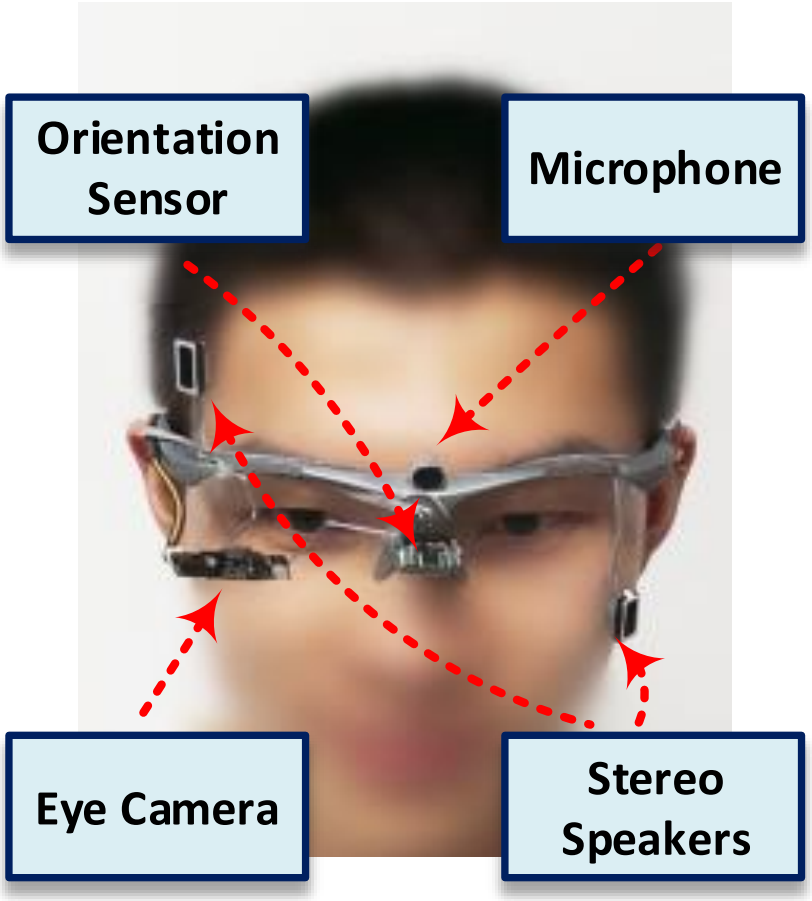
- Background
- Challenges
- System Design
- **Implementation**
- Evaluation
- Future Work

Hardware Specifications

- We design and implement a low cost smart glasses.
- We are keeping improving and upgrading our hardware design (4 different versions now)

Upgrade

Hardware Specifications



Now supports smart-phone pairing

Component	Description
Raspberry Pi II	700MHz CPU, 512MB RAM, Linux
Eye Camera	Maximum Resolution 640x480 @ 30fps
Gyroscope	MPU6050, 6-axis, Accuracy - 0.01°
Magnetic Sensor	3-axis, Accuracy - $512c/G$
Wi-Fi	OURLINK, 300Mbps, USB port
Microphone	Sample rate 44100Hz
Stereo Speakers	Frequency range 180Hz-20KHz

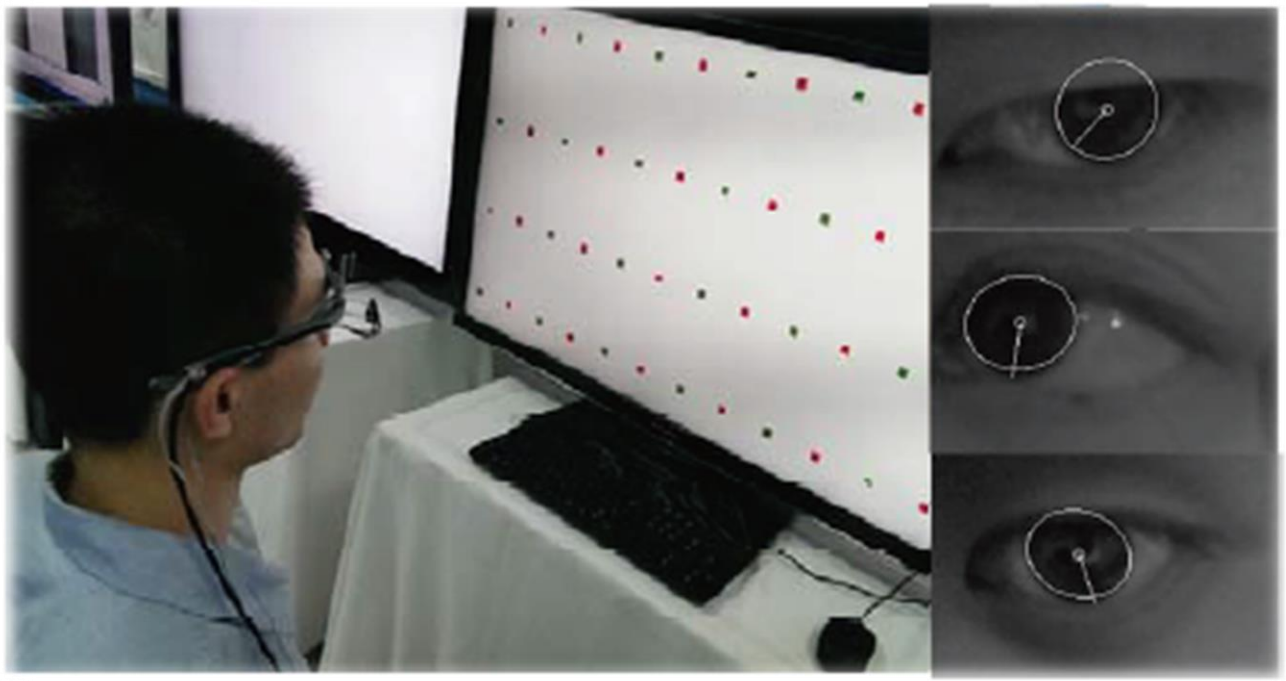
Software Specifications

- We implement two versions of all software blocks of iGaze for both **Linux** and **Android** using **C++** and **Java.**
- It supports both unidirectional and bidirectional application modes.
  - **social** application and **smart device** application
- Visual components are developed based on **OpenCV** library.
- Acoustic signal: **19kHz** and **19.5kHz**.

Outline

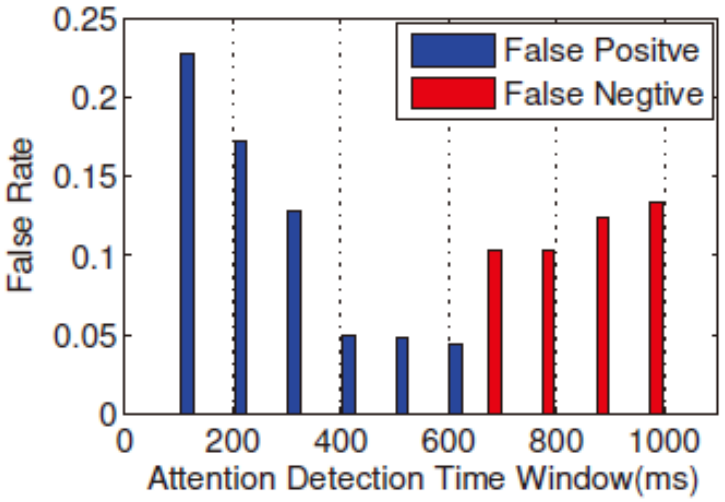
- Background
- Challenges
- System Design
- Implementation
- **Evaluation**
- Future Work

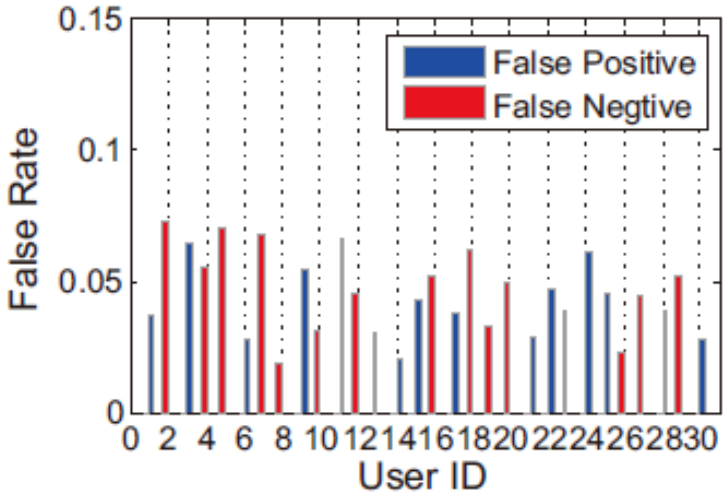
Gaze Experiment Setting



- 47 inch screen.
- Marks on the screen: 5° resolution.
- 30 Volunteers: 12 f, 18 m
- For each volunteer, about 100 gaze data are collected.

Attention & Gaze Direction Accuracy





False rate vs. Time window

- **Time-Window**
- Select **0.6s** when false positive and negative equal

False rate vs. Users

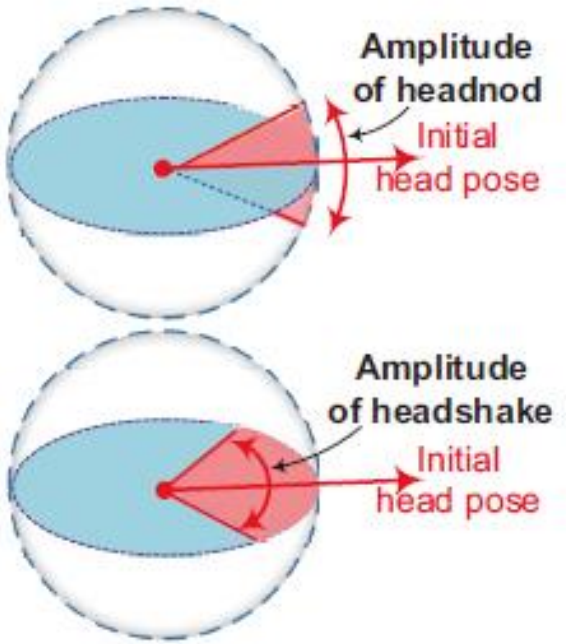
- **Eye Moving Speed**
- We set **20 pixel/s** as the **eye-moving speed** threshold.

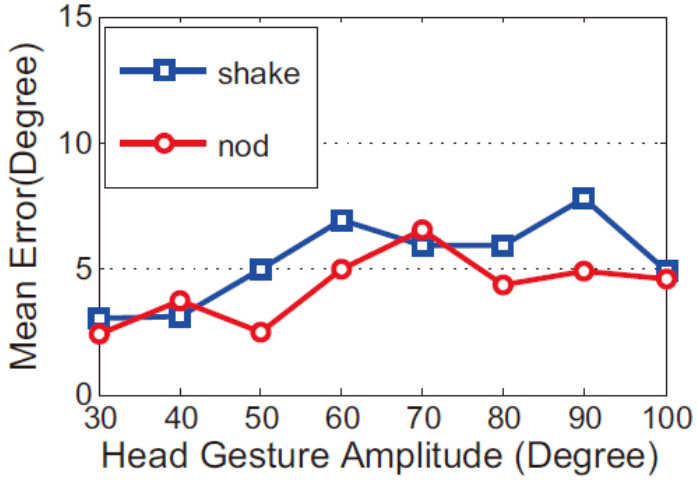
- Attention capture accuracy: **~95%**

- Gaze direction accuracy:
  - **<5° (91%)**
  - **<10° (99%)**

Device Direction Accuracy

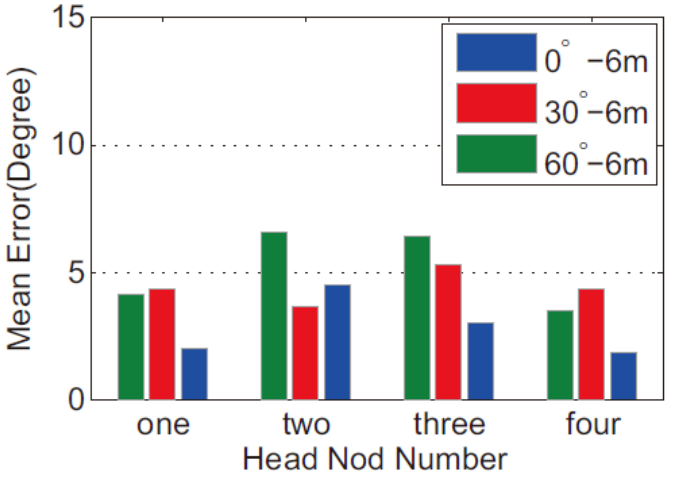
Head nod and shake





Accuracy vs. Amplitude

- High accuracy (**<4°**) can be achieved by mild head gesture (**30 °**).

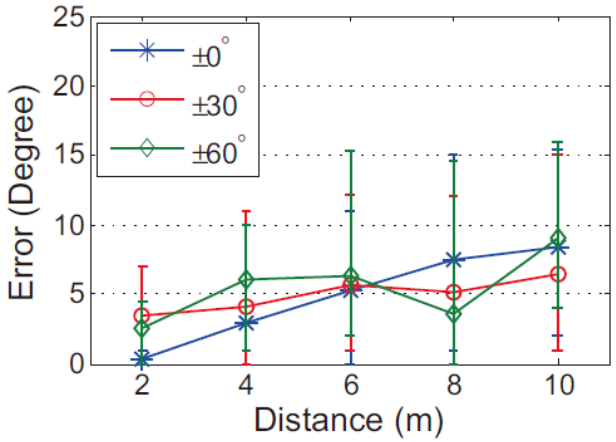


Accuracy vs. Number

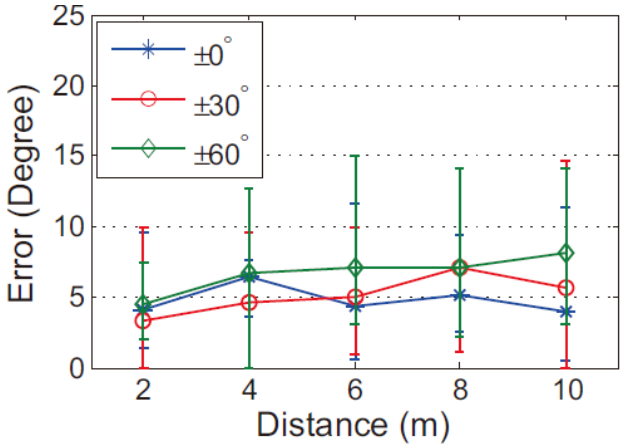
- Increasing head nod number does not improve accuracy.

- Direction estimation is robust against different head gestures and patterns.
- Only one mild head gesture is sufficient for highly accurate estimation.

Device Direction Accuracy



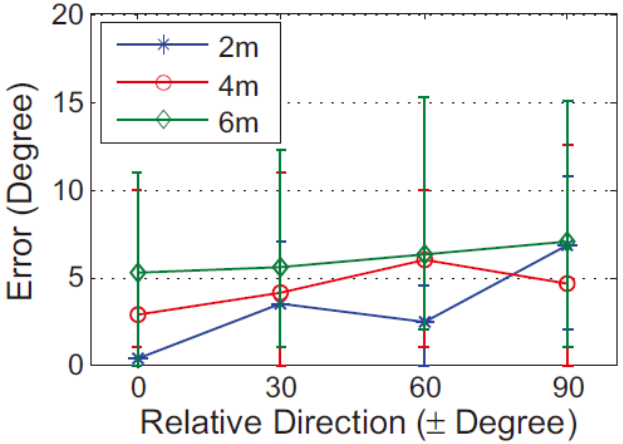
(a) Head nod.



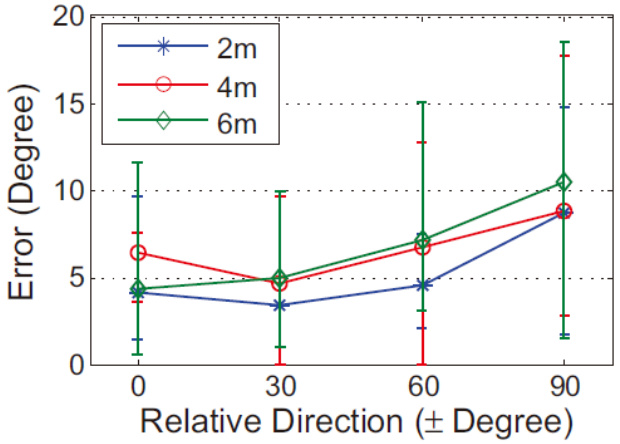
(b) Head shake.

Accuracy vs. Distance

- When the distance gets larger, the accuracy decreases.
- Error is **<5° within 4 meters** & **< 9 ° error within 10 meters**.



(a) Head nod.



(b) Head shake.

Accuracy vs. Relative angle

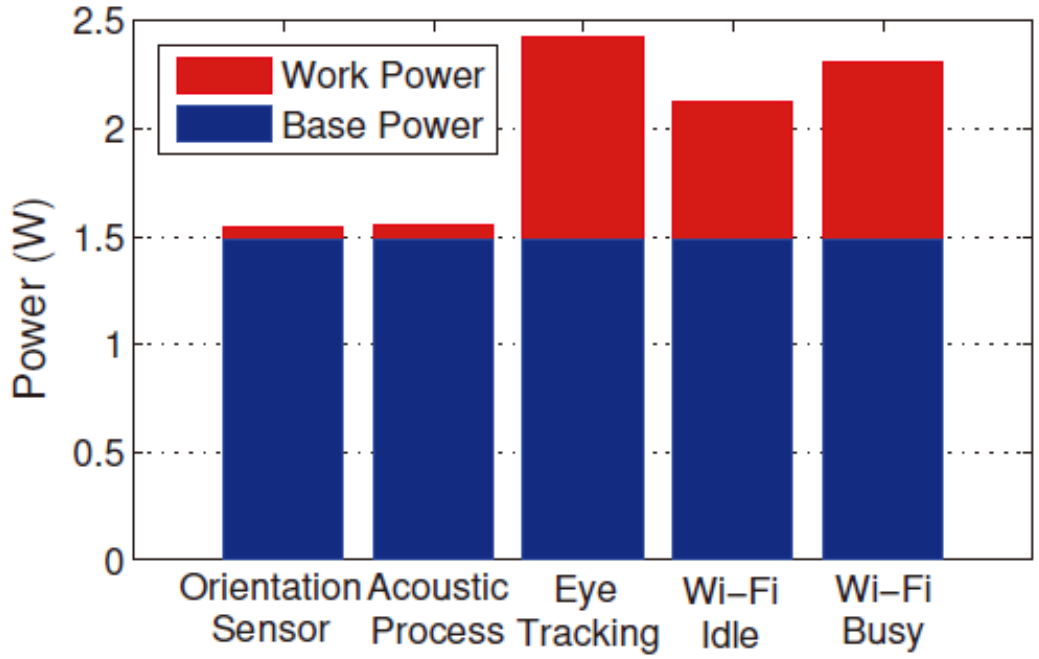
- For the head shake gesture, the mean error increases from **6° to 10°.**
- Head nodding suffers less from NLoS effect, and its mean error remains **<7°.**

Computation Cost &Energy Consumption

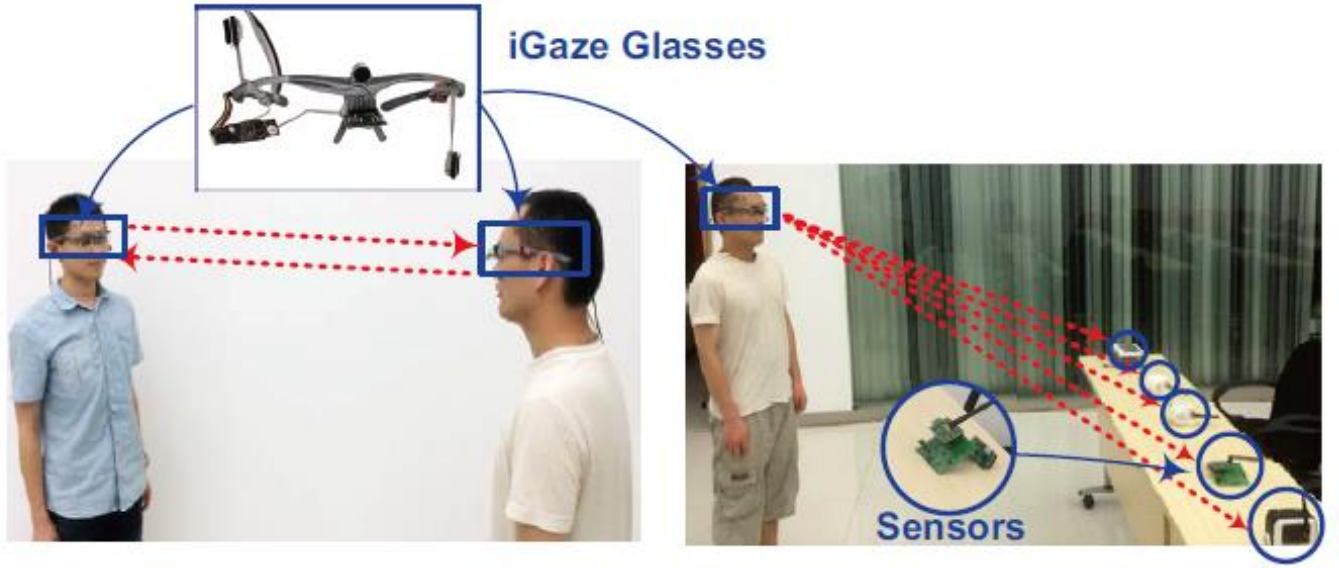
Computation Cost

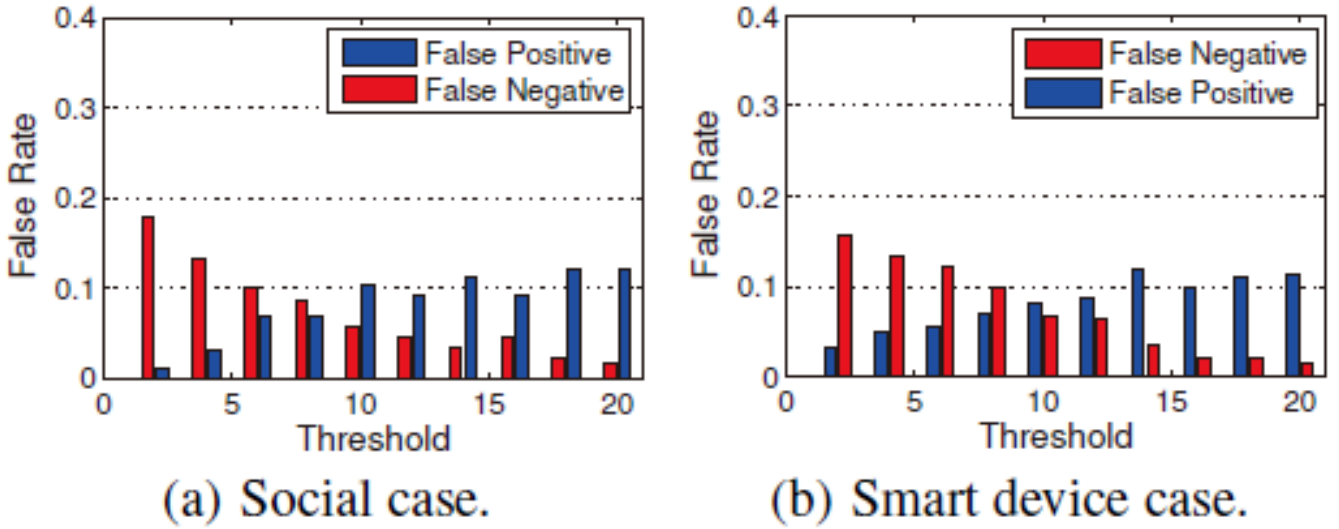
- **30 ms** to process one eye image.
- **200ms** to process the one second acoustic signal and gyro.
- Other computations are negligible.

Energy Cost



Case Study





- Considering the human factors, in most cases a connection can be built within about **3 seconds** from the initiator starts to pay attention.

- In two cases, with a proper threshold, false positive is less than **3%** while the false negative is about **15%.**

Outline

- Background
- Challenges
- System Design
- Implementation
- Evaluation
- **Future Work**

Future Work

- **Intention Capture and Understanding.**
  dynamics: multiple objects in the gaze direction, multiple users may be interested in a same object, a user may be interested in a moving object…

- **Privacy and security.**
  How to avoid insecure connection.

- **Performance improvements using a hardware-only instrumentation.**

- **Aesthetic design.**

iGaze

Thanks!

lan@greenorbs.com xli@cs.iit.edu